Detecting Single-Cell Stimulation in Recurrent Networks of Integrate-and-Fire Neurons

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik
Spezialisierung: Theoretische Physik
eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von

M. Sc. Davide Bernardi

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Benjamin Lindner (HU Berlin)
2. Prof. Dr. Igor Sokolov (HU Berlin)
3. Prof. Dr. Klaus Obermayer (TU Berlin)
Tag der mündlichen Prüfung: 6. September 2019

Zusammenfassung

Viele kognitive Funktionen des Säugetiergehirns finden im Kortex statt, der sich als großes, komplexes Netzwerk wechselseitig anregbarer Nervenzellen, Neurone genannt, beschreiben lässt. Experimentelle und theoretische Untersuchungen legen nahe, dass die präzisen Feuermuster kortikaler Netzwerke störungsempfindlich und Neurone beträchtlichen Rauschquellen ausgesetzt sind, was die verbreitete Ansicht begründet, dass nur eine große Anzahl an Neuronen als zuverlässiges Medium der Informationskodierung oder einer Verhaltensreaktion dienen kann. Experimentelle Belege dafür, dass ein einzelnes Neuron einen Einfluss auf das Verhalten eines Tieres haben kann (Houweling und Brecht, 2008; Doron et al., 2014), sind aus dieser Perspektive überraschend und sind bislang theoretisch unzureichend verstanden. Diese Arbeit ist ein erster Versuch, mit Modellbildung und mathematischer Analyse die Experimente von Houweling und Brecht (2008) zu verstehen. Diese zeigten, dass die Stimulation eines einzelnen Neurons im Barrel Cortex, der sensorische Reize aus den Tasthaaren kodiert, von einer auf die Aufgabe zuvor trainierten Ratte erfolgreich gemeldet werden kann.

Der Ausgangspunkt der Untersuchung ist ein Zufallsnetzwerk exzitatorischer und inhibitorischer Integratorneurone mit Schwellwert (*engl.* integrate-and-fire neurons) als Modell für die Umgebung der stimulierten Zelle. Dieses Netzwerk gehört zu den einfachsten Modellen, die stabile Aktivitätsmuster mit ähnlichen statistischen Eigenschaften wie im Barrel Cortex gemessen hervorbringen können. Das Experiment wird im Modell nachgebildet, indem eine aus dem Netzwerk zufällig ausgewählte Zelle stimuliert wird. Ein wichtiger Teil dieser Arbeit ist die Suche nach einem plausiblen Ausleseverfahren, das die Einzelzellstimulation mit einer mit den Experimenten vergleichbaren Zuverlässigkeit detektieren kann.

Das erste Ausleseschema betrachtet die gefilterte Aktivität einer Untermenge des Netzwerks und reagiert auf Abweichungen vom spontanen Zustand. Dieses Ausleseverfahren kann die Stimulation detektieren, wenn bei der Auswahl der Ausleseneurone denjenigen Neuronen ein ausreichender Vorzug gegeben wird, die eine direkte Verbindung von der stimulierten Zelle bekommen. Darüber hinaus ruft die Stimulation eines inhibitorischen Neurones einen größeren Effekt hervor, was mit den experimentellen Beobachtungen übereinstimmt. Diese Resultate basieren sowohl auf numerischen Simulationen als auch auf analytischen Approximationen, die verständlich machen, wie Modellparameter und Eigenschaften der spontanen Netzwerkaktivität die Detektierbarkeit der Einzelzellstimulation beeinflussen. Hier zeigt die Theorie, dass Kreuzkorrelationen zwischen einzelnen Zellen ein wichtiger Einschränkungsfaktor für die Detektion sind. Aufgrund ihrer Relevanz werden Kreuzkorrelationen auch analytisch innerhalb der Theorie der linearen Antwort untersucht.

Die mit Vorzug ausgewählte Auslesepopulation präsentiert ein erstes, einfaches Modell, wie die Tiere die experimentelle Aufgabe erlernen könnten. Es ist jedoch unrealistisch anzunehmen, dass die Entscheidung über die Ab- oder Anwesenheit des Stimulus innerhalb desselben Netzwerks, in dem auch die Stimulation stattfindet, getroffen werden kann, zumal der Barrel Cortex ein primär sensorisches Hirnareal ist. Um diese Schwäche des Modells zu beheben, wird das Ausleseverfahren im zweiten Teil dieser Arbeit erweitert, indem ein zweites

Netzwerk als Ausleseschaltkreis dient. Interessanterweise erweist sich dieses neue Ausleseverfahren nicht nur als plausibler, sondern auch als effektiver. Es benötigt nämlich einen viel schwächeren Vorzug in der Auswahl der Auslesepopulation als das andere Ausleseschema, um die gleiche Zuverlässigkeit in der Detektion zu erreichen. Der Vergleich der Simulationsergebnisse für verschiedene Konfigurationen des Auslesenetzwerks zeigt, dass die inhibitorische Vorwärtskopplung (*engl.* feed-forward inhibition) eine entscheidende Rolle spielt, da sie Inputkreuzkorrelationen größtenteils aufhebt. Ein sehr schwacher Vorzug in der Vorwärtskopplung zum Auslesenetzwerk ist notwendig, damit das Signal nicht ebenfalls unterdrückt wird. Diese Rückschlüsse werden durch analytische Rechnungen untermauert.

Weitere experimentelle Untersuchungen widmeten sich der Frage, inwiefern die Wahrscheinlichkeit der Verhaltensreaktion der Ratte von der Eigenschaften des in die Einzelzelle injizierten Stroms abhängt (Doron et al., 2014). Es wurde festgestellt, dass eine konstante Strominjektion einen Effekt auslöst, der kaum von der Dauer und Intensität der Stimulation abhängt, wohingegen eine unregelmäßige Stimulation die Reaktionswahrscheinlichkeit erheblich erhöht. Der letzte Teil dieser Arbeit befasst sich mit einer theoretischen Erklärung für diese Ergebnisse. Zu diesem Zweck wird das Netzwerkmodell erweitert, um die Eigenschaften des biologischen Systems detaillierter zu beschreiben. Zu diesen Modellergänzungen zählen unter anderem synaptische Kurzzeitplastizität, individuelle Zellparameter, realistischere Verbindungswahrscheinlichkeiten und dynamische Eigenschaften für die drei Neuronklassen, die modelliert werden (exzitatorische regulär feuernde Zellen, inhibitorische schnell feuernde Interneurone und Somatostatin exprimierende Interneurone mit niedriger Feuerschwelle). Weiterhin wird die Funktionsweise des Ausleseverfahrens so modifiziert, dass es auf Veränderungen in der Inputaktivität reagiert, anstatt wie in den vorigen Fällen den Input zu integrieren. Dieser neue "Differenzierdetektor" wird sowohl als abstrakte mathematische Operation als auch in einem Netzwerk von Neuronmodellen umgesetzt. In Übereinstimmung mit den experimentellen Ergebnissen kann der Differenzierdetektor auf die Stimulation einer exzitatorischen Zelle mit einer Zuverlässigkeit reagieren, die kaum von der Länge und Intensität einer konstanten Strominjektion abhängt, die aber bei irregulärer Stimulation zunimmt. Im Modell sind Somatostatin exprimierende Interneurone dafür entscheidend, dass exzitatorische Neurone detektiert werden können, was in der experimentellen Literatur bereits vermutet wurde. Was jedoch mit den experimentellen Beobachtungen nicht übereinstimmt, ist der relativ schwache Effekt, der sich bei der Stimulation eines inhibitorischen schnell feuernden Neurons im Modell trotz einer stark bevorzugten Auswahl der Ausleseneurone ergibt. Dies deutet darauf hin, dass der in den Experimenten meist starke Effekt inhibitorischer Neurone möglicherweise auf einem Prozess basiert, der im Modell nicht vorhanden ist. Zum Schluss wird die Hypothese diskutiert, dass ein Differenzierdetektor bei nichtstationärem Input gegenüber einem Integratordetektor vorteilhaft sein könnte.

Abstract

Many cognitive functions exerted by the mammalian brain take place in the cerebral cortex, a large, complex network of interacting excitable units called neurons. Experimental and theoretical studies suggest that the precise firing patterns of cortical networks are sensitive to perturbations and that neurons are subject to a considerable amount of noise, which justifies the belief that only large numbers of neurons can reliably carry information or elicit a behavioral response. Therefore, the accumulating experimental evidence that the activity of a single neuron can have measurable effects on the behavior of an animal (Houweling and Brecht, 2008; Doron et al., 2014) is regarded as surprising and is still theoretically poorly understood. This thesis is a first attempt at developing a theoretical model of the experiments by Houweling and Brecht (2008), which showed that a trained rat can report the stimulation of a single cell in the barrel cortex, the part of the cortex encoding tactile input from the whiskers.

As a starting point, the area surrounding the stimulated cell is modeled as a homogeneous random network of excitatory and inhibitory integrate-and-fire neurons, which is one of the most economical network models that can reproduce a stable firing pattern with the same basic statistical properties as those observed in the barrel cortex. The experiments are mimicked by stimulating one randomly selected neuron within this network. One important goal of this thesis is to seek a readout scheme that can detect the single-cell stimulation in a plausible way with a reliability compatible with the experiments.

The first readout scheme considers the filtered activity of a subset of the network and reacts to deviations from the spontaneous state. When the choice of readout neurons is sufficiently biased towards those receiving direct links from the stimulated cell, the stimulation can be detected and, as in the experiments, inhibitory cells are more easily detectable. These results are based both on numerical simulations of the network and on analytical approximations, which demonstrate how the model parameters and the properties of the network's spontaneous dynamics affect the detectability of the single-cell stimulation. In this respect, the theory shows that cross-correlations between the activity of neurons is an important factor limiting the detectability. Given their importance, cross-correlations are studied analytically in a linear-response framework.

The biased readout population is a first, simple proposal of how the animals may learn the task. However, it is perhaps unrealistic to assume that the decision about the presence or absence of the stimulus is made within the same network being stimulated, which is a primary sensory area. To overcome this limitation, the detection scheme is revisited in the second part of the thesis, in which a second network acts as the readout circuit. Interestingly, this new readout network is not only more plausible, but also more effective. More precisely, the bias needed to achieve a given effect size is much smaller than that required by the first detection scheme. The comparison of numerical simulations obtained for various readout architectures shows that feed-forward inhibition plays a crucial role by suppressing input cross-correlations. A very weak bias in the connections to the readout network is still needed

to prevent the inhibition from blocking the signal transmission as well. These conclusions are underpinned by analytical calculations.

Further experiments investigated how the rat's response probability depends on the properties of the current used to stimulate the single cell (Doron et al., 2014). Concisely, they found that a constant current injection elicited an effect that was substantially independent of the length and intensity of the stimulation, whereas an irregular current significantly increased the response probability. In an effort to explain these findings, the final part of the thesis considers a recurrent network including more biological details, such as short-term synaptic plasticity and individual cellular parameters, connection probabilities, and dynamical properties for the three neuron types included in the model (excitatory regular-spiking cells, fast-spiking inhibitory interneurons, and low-threshold spiking somatostatin-positive inhibitory cells). Importantly, the functioning principle of the readout is modified to react to changes in the activity of the local network (a differentiator readout), instead of integrating the input, as in the previous cases. This new differentiator readout is tested by using an abstract mathematical implementation and then implemented by means of a network of integrate-and-fire neurons. In agreement with the experiments, the differentiator readout can detect the stimulation of an excitatory neuron with a reliability that is essentially independent of the duration and intensity of a constant stimulus, but increases if an irregular stimulation is used. In the model, somatostatin-positive interneurons are essential for the detectability of excitatory cells, in line with earlier hypotheses from the experimental literature. However, the effect size observed upon stimulation of a fast-spiking inhibitory cell is smaller in the model than in the experiments despite a strong readout bias, thus suggesting that the large effect of fast-spiking cells observed experimentally may be due to a mechanism missing in the model. Finally, potential advantages of a differentiator readout over an integrator readout in a non-stationary situation are discussed.

Contents

Chapter 1.

Introduction

What distinguishes this sentence from a random sequence of characters? The physical processes involved in reading do not depend on the arrangement of letters. In either case, the light diffusing back from this page (or emitting from this screen) will reach the reader's retina, where light-absorbing molecules trigger a biochemical cascade of events ultimately generating a stream of electrical pulses that travel to the brain through the optic nerve. These signals eventually reach the brain region that processes visual stimuli, where electrochemical interactions cause a certain number of nerve cells, or neurons, to activate. In the end, it is the specific pattern of neurons excited in the reader's brain that determines whether a sentence like this makes sense or not.

Assigning meaning to a written text is just one of the many cognitive functions exerted by the brain, which can be seen as a huge network of interconnected neurons. The question of how a large number of interacting elements can realize abstract computations reminiscent of cognitive functions has attracted the interest of theoretical physicists for decades (Sejnowski, 1976a,b; Amari, 1977; Hopfield, 1982; Sompolinsky et al., 1988; Bialek and Zee, 1988; Treves, 1991, 1993; Amit and Brunel, 1997a; van Vreeswijk and Sompolinsky, 1998; Brunel and Wang, 2001; Monteforte and Wolf, 2010; Litwin-Kumar and Doiron, 2012; Wieland et al., 2015; Brunel, 2016; Schwalger et al., 2017). These approaches disregard the complex biophysical details of real neurons and apply concepts from statistical physics to study the dynamical properties emerging from the interaction of the single units.

In the meantime, experimental techniques to analyze neural tissues and to record the activity of the central nervous system have allowed experimental and theoretical neuroscientists to shed light on some principles underlying the function of the brain (Kandel et al., 2000; Dayan and Abbott, 2001; Gerstner et al., 2014). One approach that has proven to be fruitful for this purpose is the analysis the neural activity recorded from a brain area of an animal in a controlled situation, such as during the presentation of specific sensory stimuli or the execution of some prescribed behavioral task. This method has provided several insights, among others it allowed for the determination of the *receptive field* of many neurons, i.e. the subset of the space of all possible stimuli that elicits a response in the considered neuron. Usually, receptive fields of

neurons closer to the sensory periphery are clearly identifiable and likely to be elemental features of stimuli. In subsequent processing stages, stimulus features causing a response in a neuron tend to become more complex and abstract. A paradigmatic and well-studied example of this hierarchical nature of receptive fields is the mammalian visual system (Grill-Spector and Malach, 2004): cells in the first processing stage respond to light or dark spots in a rather precise position of the visual field; in the next stage, neurons mostly respond to a specific orientation, spatial frequency, or color; in later stages, cells can be responsive to more complex shapes irrespective of their position in the visual field, to the speed of motion in a particular direction, and even to objects or to a specific famous actress (Quiroga et al., 2005).

If neuronal responses become increasingly selective, it could be expected that the neural code becomes increasingly sparse, as it was proposed decades ago (Barlow, 1972). In other words, the representation of complex percepts would be based on a small number of neurons active at a given time. It has been argued that this *sparse coding* may bring several advantages, ranging from the maximization of information capacity to the optimization of energy usage, as neuronal discharge has a high metabolic cost (Földiak, 1990; Olshausen and Field, 2004). Despite the possible benefits, a code relying on a very small number of active neurons bears risks in terms of reliability, as most biophysical processes governing the generation and transmission of *spikes*, the neuronal discharges used by neurons to interact with each other, are stochastic and often quite unreliable. Furthermore, theoretical studies suggest that the spiking pattern of recurrent networks can be very sensitive to perturbations (van Vreeswijk and Sompolinsky, 1996, 1998; Monteforte and Wolf, 2010), as also supported by some experimental evidence (London et al., 2010). Strong noise and correlations between neurons led to the assertion that only a large population of cells can support a functioning code (Averbeck et al., 2006). Whether the strategy employed by the brain to encode complex concepts is based on a representation distributed across many neurons (a population code), on few neurons (a sparse code), or even on single cells (so-called grandmother neurons), is still an open question (Gross, 2002; Wolfe et al., 2010; Rolls and Deco, 2010; Barth and Poulet, 2012; Rolls, 2017). On the contrary, it is widely accepted that, in mammals, the location in which many complex cognitive processes take place is the cerebral cortex (Kandel et al., 2000).

The cortex is the large brain region where input from different sensory systems converge, where they are combined with each other and with past experiences, where decisions are formed, and from which motor control originates (Kandel et al., 2000). Because different areas of the cortex are both specialized and heavily interconnected, it is difficult to attribute recorded neuronal activity patterns to a defined external factor. To circumvent this problem, the approach of "reverse physiology" employs direct manipulation of the cortical activity to influence behavior (Brecht et al., 2004a; Doron and Brecht, 2015). Cortical microstimulation, i.e. the injection of brief current pulses into the extracellular space, has long been used to induce activity in

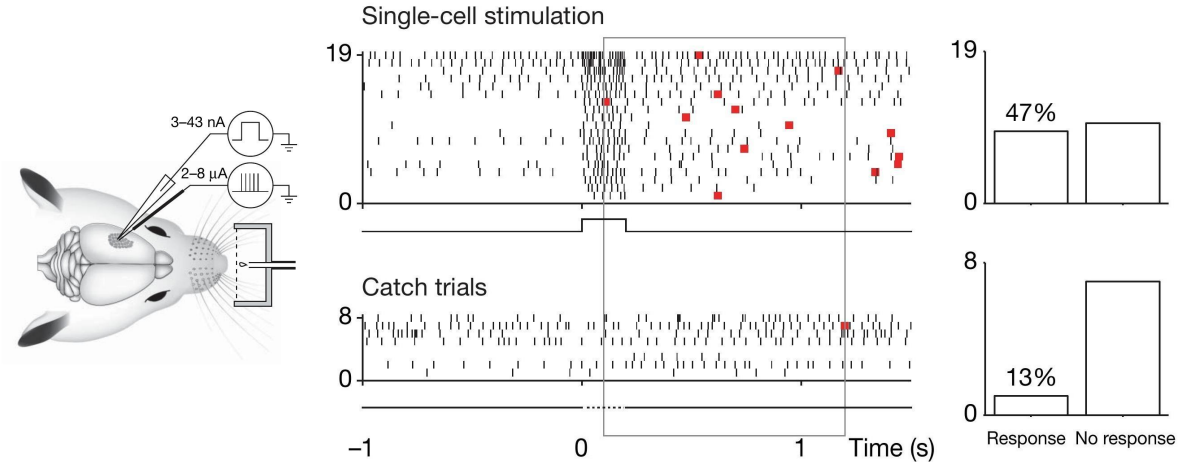

Figure 1.1. – **Rats can be trained to report stimulation of a single cortical cell.** Stimulation of this neuron was reported 47% of the times (red squares), while false positives (catch trials) occurred in 13% of the trials (adapted with permission of Springer Nature from Houweling and Brecht, 2008).

localized neuronal populations (Asanuma and Sakata, 1967). Analyzing the behavioral responses induced during controlled tasks enabled the understanding of the role of specific brain areas in the formation of sensation and behavior (Salzman et al., 1990; Salzman and Newsome, 1994; Tehovnik, 1996). In recent years, the selectivity of cortical stimulation was improved to the point that reliable manipulation of the activity of a single neuron in the central nervous system *in vivo* is possible (Houweling et al., 2010). By employing this technique, several experiments came to the surprising conclusion that single-cell stimulation in the cortex of a living organism can produce a measurable effect on the local network activity (Bonifazi et al., 2009; Li et al., 2009; Kwan and Dan, 2012) or even influence the behavior of the animal (Brecht et al., 2004b; Houweling and Brecht, 2008; Doron et al., 2014; Tanke et al., 2018).

In particular, Houweling and Brecht (2008) have shown that rats previously trained to report microstimulation can be biased to respond to the stimulation of a single neuron in the barrel cortex, the part of cortex encoding tactile input from the rat's whiskers. For instance, the cell stimulated in fig. 1.1 increased the rat's response rate from 13% (false positive rate) to 47%. This increase is rather large but not statistically significant because of the limited number of trials. The average effect over many cells is weaker, but indeed statistically significant (more details on the experiment are given in section 1.2). Considering that the rat cortex contains more than 20 million neurons (Korbo et al., 1990; Herculano-Houzel et al., 2011), of which more than half a million form the barrel cortex, the idea that few extra spikes induced in one cell can make a difference may surprise even a supporter of the sparse coding hypothesis.

One simulation study exists that examined the effect of stimulating one cell in a small bursting

network (Luccioli et al., 2014). Other previous numerical studies investigated the trade-off between stability and sensitivity with respect to the repeated stimulation of few cells (mimicking microstimulation) in networks of different topologies (Vasquez et al., 2013; Martens et al., 2017). However, a theoretical model of how the simulation of a single cell can be detected in a large network, as in the experiment by Houweling and Brecht (2008), is still missing. The aim of this thesis is to attempt a first step in this direction. Several factors render the theoretical modeling of the experiments by Brecht and coworkers quite challenging: the daunting complexity of the system, the unknown effects of the training phase the animals undergo, and the unclear link between the neuronal activity of the network and the behavioral response. The guiding principle to approach these three difficulties will be to try to simplify the description as much as possible.

The level of description adopted here to model neuronal activity is that of recurrent networks of integrate-and-fire point neurons, introduced in detail later in this chapter. Although they are a simplified phenomenological description of the neuronal dynamics, integrate-and-fire neurons can reproduce the firing pattern of some cortical neurons with reasonable approximation (Badel et al., 2008; Gerstner and Naud, 2009). Furthermore, random networks of integrate-and-fire are one of the most economical recurrent network models that can reproduce basic spiking patterns observed in recordings from cortical networks, such as desynchronized, irregular firing, and oscillations on various time scales (Brunel, 2000). In these models, chaotic noise-like fluctuations arise naturally from the combination of spiking dynamics and the random interactions (van Vreeswijk and Sompolinsky, 1996). In other words, the noise emerges as an intrinsic property of the system and it does not need to be added *ad hoc*. This feature of the model is quite relevant when attempting to read out the effect of a weak signal against the spontaneous background activity of the network in which the stimulated neuron is embedded.

As mentioned above, the ability to report the single-cell stimulation required a training phase, i.e. a directed modification of the system. Specifically, the training phase made use of microstimulation pulse trains, the injection of current pulses in the µA range into the cortical extracellular space. In contrast to the single-cell stimulation, the effects of cortical microstimulation are difficult to assess with precision because it impacts quite unspecifically a complex system, triggering multiple effects on different time scales. Although two main effects have been consistently identified, namely the direct activation of a sparse pattern of neurons within a radius of about one mm (Histed et al., 2009) and a long-lasting ($\sim 100\,\mathrm{ms}$) inhibition within a similar radius (Butovas and Schwarz, 2003; Butovas et al., 2006), computational modeling of cortical microstimulation suggests that the precise effects on the network depend on many details, such as the exact depth of stimulation, the intensity of the current, and the relative position, type, and orientation of neurons in the surroundings of the electrode (Overstreet et al., 2013). Given the large number of unknown factors, in this thesis the training phase will not be explicitly modeled, and the system will be assumed to be in the state after the learning process has taken place. A possible

way of quantifying the extent of the modification from the naive state will be introduced.

Analytical solutions for the dynamics of coupled integrate-and-fire neurons are generally not easy to obtain. Therefore, numerical simulations will be a primary tool in this thesis. However, analytical approximations will be sought whenever possible as a valuable instrument to gain insight into simulation results.

Outline of the thesis

The remainder of this chapter offers a brief review of concepts used in the following chapters. Section 1.1 focuses on the basic biological notions about neuron physiology and elementary facts concerning cortical networks, with particular emphasis on the rat barrel cortex. Because of their centrality in this thesis, the experiments by Brecht and coworkers are described in more detail in section 1.2. The mathematical definitions used in this thesis to describe the neural activity are introduced in section 1.3. Section 1.4 deals with the mathematical models of neurons and synapses employed in the following chapters. Finally, section 1.5 introduces some basic notions on the mathematical description of networks and briefly describes the Amit-Brunel network model, which is used as foundation for the models considered in this thesis.

Chapter 2 deals with a proof of principle. The core idea is to consider one of the simplest spiking network models that autonomously generates a firing pattern compatible with the spontaneous activity in the barrel cortex, to stimulate a randomly selected neuron, and to determine if and under what circumstances a suitable readout scheme can detect the occurrence of the stimulation. The main result is that a readout scheme receiving input from a sufficiently biased selection of neurons in the network can detect the stimulation. It can be hypothesized that this readout bias results from the training.

Chapter 3 introduces a second network acting as a readout circuit. Remarkably, this addition not only increases the plausibility of the readout mechanism, but even improves its effectiveness.

Chapter 4 attempts to explain the results obtained from a second series of experiments investigating how the parameters of the single-cell stimulation modulate the strength of the behavioral response (Doron, 2012; Doron et al., 2014). To this end, a network model endowed with more biological details is considered. If the readout mechanism introduced in chapter 3 is modified to operate as a differentiator circuit, several properties of the experimental data can be captured.

Some of these results have been published elsewhere, namely the core results of chapter 2 (Bernardi and Lindner, 2017) and the results of chapter 3 (Bernardi and Lindner, 2019). The software developed for the network simulations of this thesis was used to perform network simulations of two further papers (Wieland et al., 2015; Pena et al., 2018).

1.1. Basic biological background

This section briefly reviews the most basic facts about the morphology and functionality of neurons, of their interactions, and of cortical networks. Whenever an explicit literature reference is missing, the reader can refer, for instance, to the textbook by Kandel et al. (2000).

Neurons, ions, action potentials

The external membrane of neurons is to a large extent impermeable to most ions. However, specific ions can cross the membrane owing to the presence of specialized proteins. Some of them act as ion pumps that actively transport specific ion types across the membrane, even against a potential difference or a chemical gradient. In neurons, ion pumps maintain an imbalance in the concentration of several ion classes across the membrane, so that an electric potential difference results. This *resting potential* is typically in the range $-80\,\mathrm{mV}$ to $-70\,\mathrm{mV}$, if voltages are measured with respect to the extracellular space. Another important set of membrane proteins are *ion channels*, capable of letting only specific ion types flow though the membrane. Some of these channels may have only one configuration, i.e. they act as selective, but static gates. Many channels, however, can switch between an open and a closed state. These transitions between channel states are stochastic but the transition probabilities for the one or other direction are heavily influenced by external causes, such as the potential difference across the membrane (for voltage-gated channels) or the binding of specific molecules to the channel's receptor (ligand-gated channels). Ionic currents regulated by voltage-gated channels are the basis for the formation and propagation of the electrical signals used by neurons to interact with one another, as described below.

Morphologically, neurons can be very complex (see fig. 1.2 for a simplified sketch with labels and 1.3A for a reconstruction of some actual neurons). On a coarse level of description, three morphological components can be distinguished: the main cell body, called *soma*, and two complex tree-like structures stemming from it: the *dendrites* and the *axon*. The soma is the metabolic center of the cell, it integrates input from other cells, and the somatic membrane potential is the main factor initiating the response of the neuron. In a simplified picture, the main function of dendrites is to collect input signals from other neurons and to relay it to the soma, although current research is uncovering an active role of dendrites in the neural computation (see for instance Larkum, 2013).

Inputs from other cells affect the membrane potential in the soma. When the somatic membrane potential reaches a certain threshold value (typically $10\,\mathrm{mV}$ to $30\,\mathrm{mV}$ above the resting potential, depending on the cell type), the concerted activation of voltage-gated sodium and potassium ion channels generates an *action potential* (AP), the main electrical signal used by neurons to communicate. The shape of APs is stereotypical and consists of a rapid depolariza-

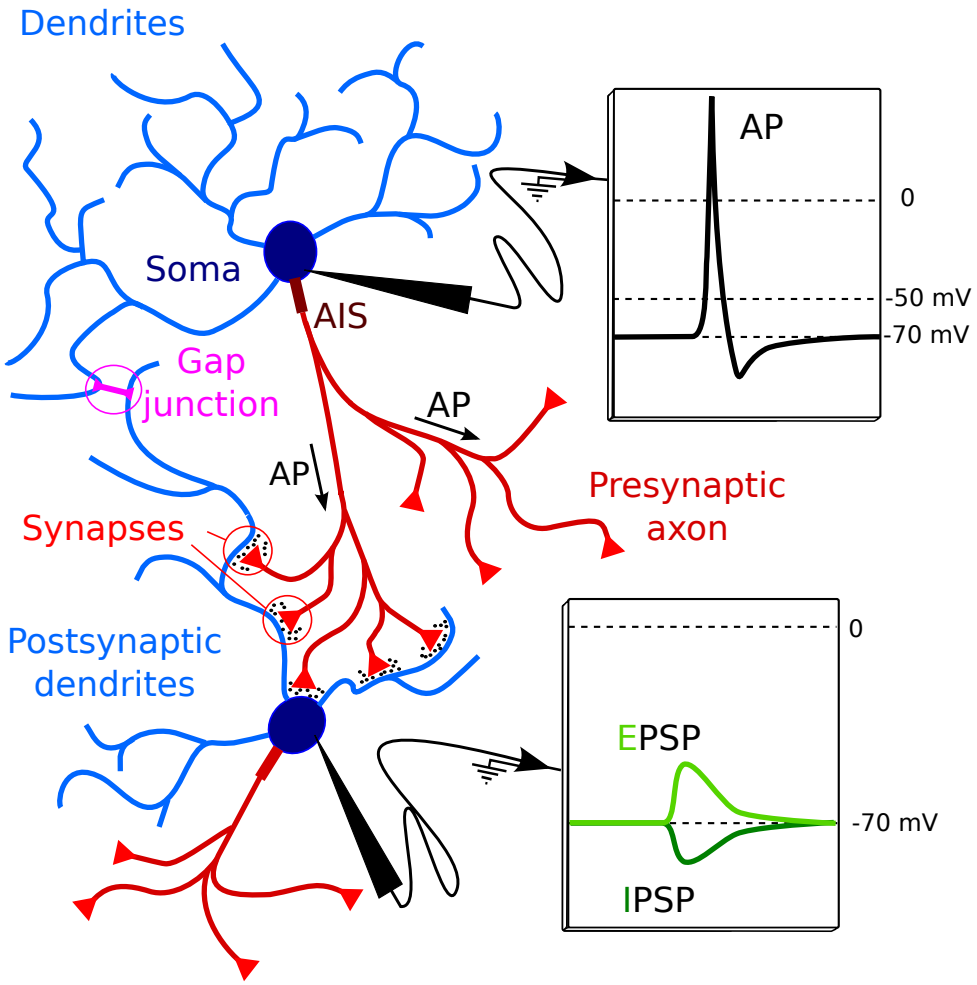

Figure 1.2. – **Sketch of a neuron and of synaptic transmission**

tion upswing followed by a fast hyperpolarizing downswing. The typical amplitude of one AP is of about 100 mV and its width is usually of 0.5 ms to 2 ms, depending on the neuron type.

Synapses, plasticity, connection patterns

Action potentials, also commonly referred to as *spikes*, are usually first triggered in the axon initial segment (AIS), where the active-channel density is highest, and then propagate along the axonal tree until they reach the axon terminals. Axonal terminals in the close proximity (20 nm to 40 nm) of the soma or dendrites of a second neuron can form a *synapse*, which allows the signal carried by the AP in the first neuron (termed *presynaptic*) to propagate to the second neuron (defined *postsynaptic*) *unidirectionally* (Cowan et al., 2003). When the AP arrives at a synaptic terminal, specific molecules, called *neurotransmitters*, are released into the *synaptic cleft*, the narrow gap between the axon terminal and the membrane of the postsynaptic cell.

Neurotransmitters can bind to specific ion channels in the cell membrane of the postsynaptic cell, causing them to open. When these channels open, the conductance across the postsynaptic cell membrane changes, thus leading to a postsynaptic potential (PSP), a transient change in the membrane potential of the postsynaptic cell. Depending on the neurotransmitter and ion channel type, the current can be depolarizing or hyperpolarizing. In the former case the presynaptic cell has an *excitatory* effect, i.e. it transiently increases the voltage within the postsynaptic cell by depolarizing it and the resulting PSP is also said excitatory (EPSP). In the other case, a hyperpolarizing current generates an *inhibitory* PSP (IPSP), that is, a transient trough in the postsynaptic potential. The terms excitatory and inhibitory are due to the fact that EPSPs increase the probability that the postsynaptic cell fires, whereas IPSPs makes a spike in the postsynaptic cell less likely.

Early studies on synaptic transmission suggested that all axon terminals belonging to the same neuron release the same kind of neurotransmitter, and that, therefore, each neuron has the same effect on all its postsynaptic targets, a fact that became established as *Dale's principle* (Strata and Harvey, 1999). Although it is now known that the same neuron can produce multiple neurotransmitters and that Dale's principle is no universal law (Jonas et al., 1998; Cowan et al., 2003), most neurons can be labeled as either excitatory or inhibitory, depending on the neurotransmitter type(s) released by its outgoing synapses.

The processes underlying the functioning of synapses are stochastic (Keener and Sneyd, 2009). As a consequence, the output of each synapse can be quite variable and each PSP exhibits a certain degree of randomness in size and reliability. As a matter of fact, transmission failures are not uncommon in most synapses. While strong synapses are usually found to be more reliable, in some cases failure rates can be as high as 60% (Beierlein et al., 2003; Helmstaedter et al., 2008).

Another important feature of synapses is that their strength is not constant in time, but depends on the past activity. This phenomenon is known as synaptic *plasticity* and can happen over very different time scales (Cowan et al., 2003). *Short-term plasticity* (STP) occurs over time scales ranging from a few ms to few seconds and is related to the cellular mechanisms of neurotransmitter release and uptake. The effect of STP is that repeated stimulation of a presynaptic cell leads to PSPs that can decrease or increase in amplitude (fig. 1.3B). The former case is known as *short-term depression* and the latter as *short-term facilitation*. Sometimes, a synapse can show both effects depending on the firing rate of the presynaptic cell, i.e. the number of APs emitted by the presynaptic cell per unit time. Effects of STP usually wear off after some seconds, at most. However, the firing activity of neurons can lead to a long-lasting alteration of the synaptic strength. This long-term plasticity happens over time scales ranging from hours to days, and is known as long-term potentiation, if the synaptic strength is enhanced, or long-term depression, when the synaptic strength is weakened. Long-term plasticity is believed to be one

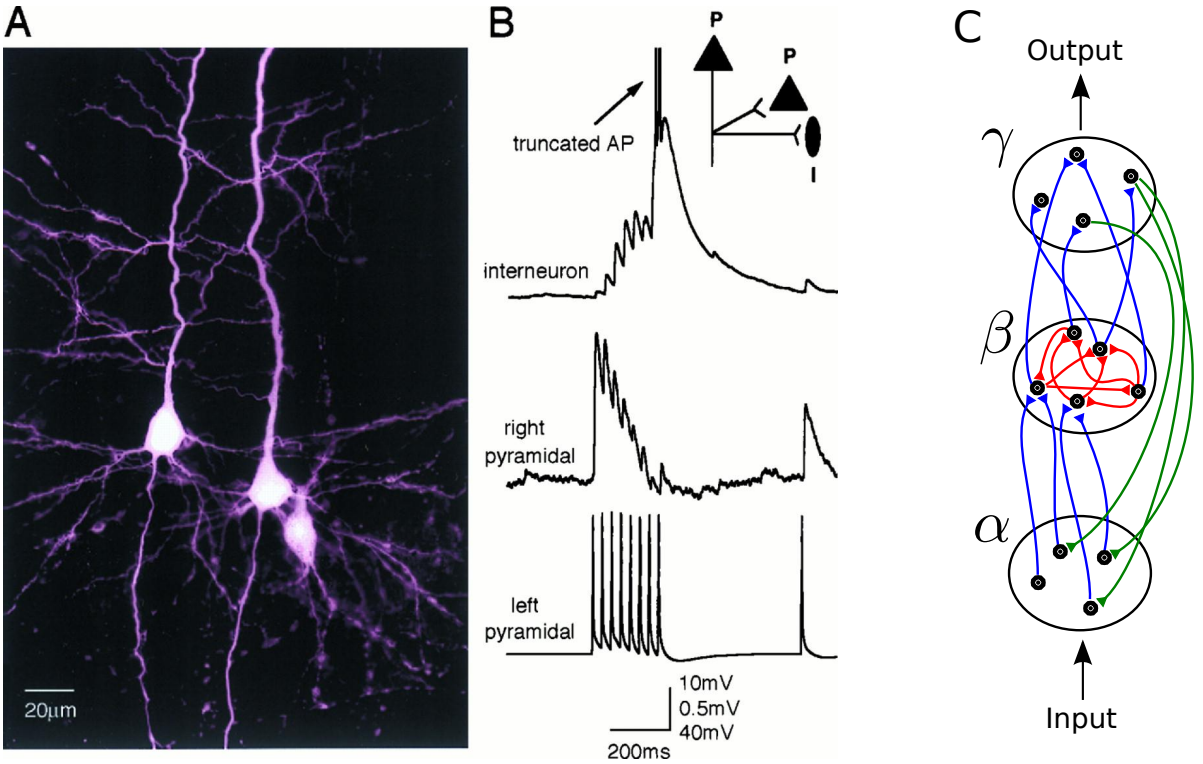

Figure 1.3. – **A**: Reconstruction of two pyramidal cells and one interneuron in the rat cortex. **B**: Short-term plasticity for the synapses between the three neurons in **A**. The left pyramidal cell is stimulated and generates a train of action potentials (bottom trace). The postsynaptic potentials recorded in the right pyramidal cell show short-term depression (middle trace), while those recorded in the interneuron (top trace) display short-term facilitation and lead to the generation of an action potential. **A** and **B** are adapted from Markram et al. (1998), Copyright (1998) National Academy of Sciences. **C**: Connection patterns between interconnected populations. Blue: feed-forward connections, Red: lateral recurrent connections. Green: feed-back connections.

main mechanism underlying learning processes in the brain.

Because their action is mediated by the diffusion of molecules, the synapses described until this point are also called chemical synapses, as opposed to another type of connections between neurons, in which the intracellular space of the two neurons is physically connected by specific channels. These channels are known as *electrical synapses* because they permit a direct current flow between two neurons. Electrical synapses are also called *gap junctions* and usually connect dendritic trees of cells of the same class (Gibson et al., 1999; Galarreta and Hestrin, 1999). Although gap junctions are common in some areas, they are found only in specific cell classes, while chemical synapses are the most common type of synapse found in the central nervous system. Therefore, the term synapse without adjective is often meant to indicate a chemical synapse. In the following, the term synapse will only indicate a chemical synapse, while electrical synapses will be termed gap junctions.

The wiring diagram of the central nervous system, sometimes called the *connectome* by analogy with the genome, is a key factor in the functioning of the brain. On the fine scale, there is often a substantial degree of randomness in the connection patterns between single neurons. However, on larger scales, the connectivity exhibits a structure, in which neurons can be grouped on the basis of anatomy and function. The term neuronal *population* is sometimes used to describe an aggregation of neurons defined by spatial proximity, morphology, and function. Although complex connection diagrams are cumbersome to capture in words, there is a terminology to describe basic connection patterns involving two or more populations. Suppose that the three populations schematically represented in fig. 1.3C perform a generic computation by receiving input from some sensory system, which influences the activity of the population α, and by providing an output response based on the activity of the population γ. All connections depicted in blue are termed *feed-forward*: they connect α to β and β to γ in a unidirectional way. If the indirect effect through the green *feedback* projections from γ to α is ignored, neurons in α do not influence each other. On the contrary, neurons within β can affect one another via the *recurrent* connections painted in red. The classification of the connections in feed-forward, feedback, recurrent can be context-dependent. For instance, the distinction between feed-forward and feedback is possible in the example of fig. 1.3C only because of the functional role assigned to α and γ, but is not a property of the connections themselves. Furthermore, it is clear that the three populations considered as a whole form a recurrent network, because neurons in α can influence each other though β and γ, and the same applies to those in γ. However, the above classification of the connections still makes sense when the three populations are seen as distinct steps in a sequential operation. Local recurrent connections such as the red ones within β are sometimes called *lateral* to emphasize that they are restricted to a single processing stage.

Neocortex, whiskers, and the rat barrel cortex

Neural populations can be as small as a handful of neurons or comprise several thousands of cells. Even anatomically small brain regions usually contain a very large number of neurons, and thus multiple populations. One remarkable and quite large anatomical region of the brain is the neocortex, a late product of evolution distinctive to the mammalian brain (Rakic, 2009). Anatomically, the neocortex is the outer layer of neurons enclosing the cerebral hemispheres. It is located right under the skull, and its depth is in the range 0.5 mm to 4 mm depending on the species and on the cortical area (Rockel et al., 1980). Because of its extension, the neocortex has been divided in a large number of anatomical subdivisions (Paxinos and Watson, 2006). While its overall size and anatomical details can differ across different mammalian species (for instance the characteristic wrinkled surface of the human cortex is absent in rodents, Semple et al., 2013), several organizational and structural principles are conserved both across species and cortical areas. One prominent characteristic shared by different cortical regions and animals

is the layered structure. The traditional classification identifies six layers in the majority of the cortex, layer I being the most superficial, and layer VI the deepest one.

Layer IV is usually considered the input layer, as most connections from sensory pathways terminate here. A typical source of these input connections is the thalamus, the brain region located under the neocortex that is one stage closer to the sensory periphery. Layer IV is typically endowed with recurrent connections and ascending vertical projections to the upper layers II/III, from which outgoing connections extend both horizontally to other cortical areas and vertically to local deep layers (below layer IV). Feedback projections to brain areas upstream in the sensory processing (such as the thalamus) have mostly been found to originate from deep layers. These canonical roles of the different layers have been identified by comparing recurring vertical connection patterns in different areas and species (Douglas and Martin, 2004). These and further similarities led to the hypothesis that the cortex is, to some extent, modular and that it may have evolved by adapting to different contexts and tasks one elemental microcircuit, called cortical column because of its vertical arrangement (Mountcastle, 1997). The idea of a standardized construction element is appealing because it would imply that generic computational principles applicable to all cortical areas may exist (Douglas and Martin, 2004, 2007a,b). However, to what extent the concept of a cortical column is for real it is still debated (da Costa and Martin, 2010).

Cortical neurons have a great variety of shapes, sizes, and functions. One major possible classification is that between local interneurons and projection neurons. *Local interneurons* form connections locally and are mostly inhibitory. *Projection neurons* are mostly excitatory have often a pyramidal-shaped cell body, and are thus often called *pyramidal cells*. Pyramidal cells usually have large dendritic trees that spread preferentially in the vertical direction, and can extend across all layers. Their axons can form contacts both with nearby and distant cells.

Functionally, the neocortex is believed to be the neural basis of most complex cognitive functions such as the integration of sensory information of different kinds, decision making, the planning and execution of motor outputs. Although most behavioral tasks engage more than one cortical area, different regions of the cortex are specialized. An ongoing effort of neuroscience is the development of *cortical maps* that link the anatomical cortical area to its function(s).

One example is given in fig. 1.4A, which shows an illustration of the rat's brain (reproduced from Feldmeyer et al., 2013), in which some cortical areas have been labeled and marked by color shadings. The large part of the sensory cortex painted in red is related to somatosensation, i.e. the processing of tactile stimuli. The area marked in light blue, the prefrontal-premotor (PFPM) cortex is associated to decision making and cognitive control, while the green shading indicates areas pertinent to motor control. To the right of the somatosensory cortex, the auditory (Aud) and the visual (Vis) cortex are also labeled.

One of the most studied model systems in neuroscience is the part of the rat brain processing

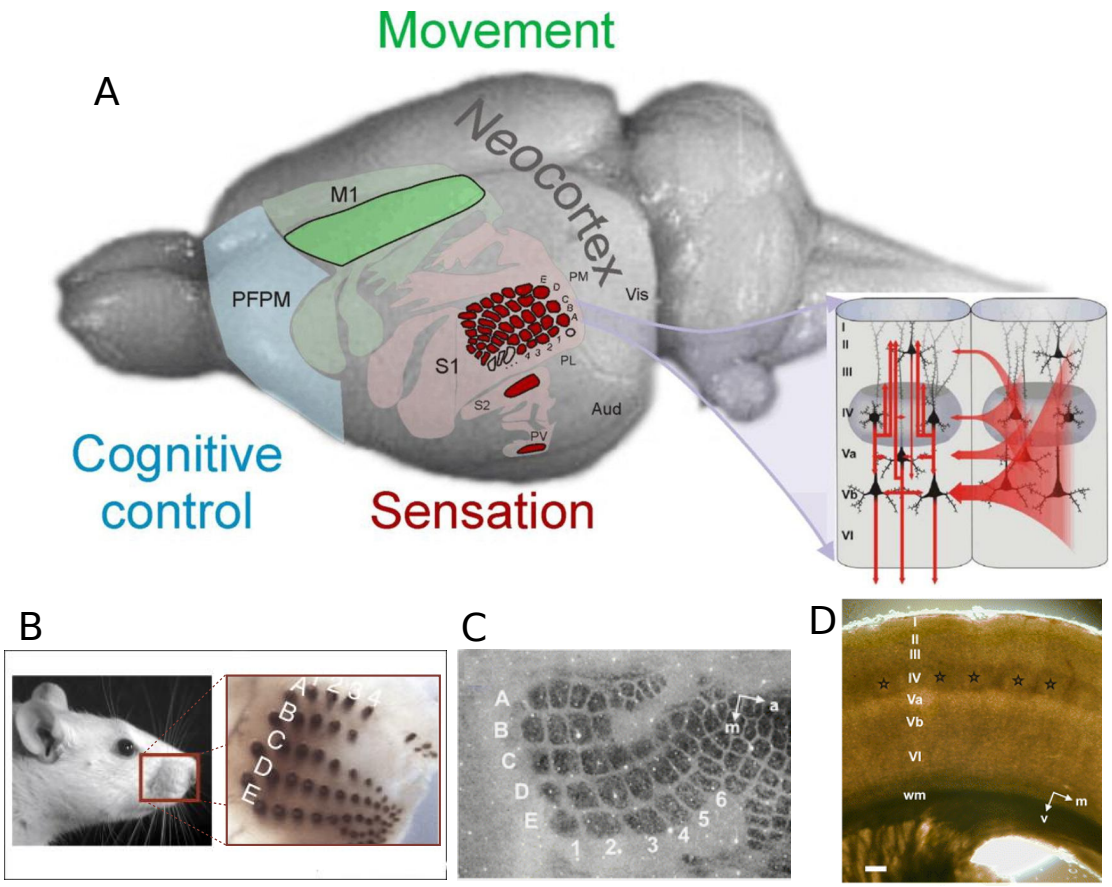

Figure 1.4. – **A**: Illustration of the rat brain adapted from Feldmeyer et al. (2013) by permission of Elsevier. Some cortical areas are labeled, and areas related to the whisker system are marked with different colors. The inset, originally taken from Schubert et al. (2007), represents the main vertical excitatory connections within a barrel column. **B**: Picture of the rat whiskers with close-up on the whisker pad. Whiskers are organized in rows (labeled by letters) and arcs (labeled by numbers). Adapted with permission of Elsevier from Knutsen and Ahissar (2009). **C**: Cytochrome oxidase staining of a tangential section through layer IV of the rat barrel cortex (seen from above). The spatial arrangement of barrels reproduces that of whiskers. **D**: Acute slice through the barrel cortex of a rat. Layers are indicated by Roman numerals and barrels are marked by stars. **C** and **D** were adapted with permission of Springer Nature from Schubert et al. (2007).

whisker-related information (Feldmeyer et al., 2013). Rats rely strongly on their whiskers to construct a sensory representation of the external world and to navigate it (Vincent, 1912; Diamond et al., 2008). Accordingly, quite large portions of the primary sensory (S1) and primary motor cortex (M1) were found to be involved with the whisker representation. These areas and are highlighted in dark green and red in fig. 1.4A. The word *primary*, referred to sensory cortical areas, indicates those parts of the cortex that directly receive input connections from sensory nerves, while it indicates regions that directly provide output to motor neurons, when referred to motor cortical areas. Among these highlighted areas, the one belonging to S1 is known as *barrel cortex* and is of particular significance to this thesis, because it is the cortical area in which the single-cell stimulation experiments were conducted.

The barrel cortex owes its name to one distinctive functional and, partially, anatomical feature: the layer IV is organized in barrel-shaped areas that encode sensory information from single whiskers in a topographical representation (Woolsey and Van der Loos, 1970). There is a one-to-one correspondence between whiskers and barrels, in which even the spatial arrangement in rows and columns of whiskers on the body is conserved in the cortex (compare the close-up of the rat's whisker pad in fig. 1.4B to the top view of the barrel cortex in fig. 1.4C). The anatomical separation of barrels is present in two aspects: the input connections that originate from the thalamus and that terminate mostly in layer IV; and in the local recurrent connections within layer IV (fig. 1.4D). The space in-between barrels is called septum. When a single whisker is stimulated, the initial response is prevalently confined to the corresponding barrel in layer IV (Brecht and Sakmann, 2002). Afterwards, the excitation propagates to upper layers, where it spreads to surrounding columns (Petersen et al., 2003).

Larger barrels have a diameter of about 300 µm. The cortical column enclosing them is 1.7 mm to 1.8 mm high and it contains roughly 19 000 neurons (Meyer et al., 2010), of which 10 % to 20 % are inhibitory interneurons (Meyer et al., 2011). Connections between excitatory neurons are mostly sparse (with a connection probability of 5 % to 10 %), organized along vertical and horizontal axes. Vertical connections are mostly layer specific and display many of the features of the canonical cortical microcircuit (Lefort et al., 2009; Feldmeyer, 2012; Avermann et al., 2012). Inhibitory connections are denser, with connections probabilities as high as 40 % to 70 % for local connections (Gibson et al., 1999; Beierlein et al., 2003; Packer and Yuste, 2011; Avermann et al., 2012; Koelbl et al., 2015). Horizontal excitatory axons within the barrel cortex target either the surrounding septum cells or neighboring barrels, with a preference to extend along whisker rows (indicated by letters in fig. 1.4) rather than arcs (Petersen et al., 2003). Long-range horizontal connections originating from barrels terminate predominantly in the secondary somatosensory area S2, while long-range axons stemming from septum cells target mainly the primary motor cortex M1 (Feldmeyer et al., 2013).

1.2. Single-cell stimulation experiments in the barrel cortex

The experiments by Brecht and coworkers showing that the behavior of an awake rat can be influenced by stimulating a single cell in the barrel cortex have been published in two papers (Houweling and Brecht, 2008; Doron et al., 2014). Because of their importance for the models of this thesis, these experiments are briefly described in this section.

First behavioral report of single-cell stimulation (Houweling and Brecht, 2008)

The experimental setup is sketched in fig. 1.5A. Head-fixated rats were trained to report short (200 ms) trains of current pulses in the µA range delivered to the extracellular space in deep cortical layers within the rat barrel cortex. Rats were awake and were trained to respond by interrupting a light beam with their tongue.

After rats learned to reliably report microstimulation at low intensity (in the range 2 µA to 5 µA), a single cell in the vicinity of the electrode used for microstimulation was approached with a glass pipette, which was used to inject current into the selected cell without penetrating the cell membrane, which is perforated by the current. This technique, originally developed for cell staining (Pinault, 1994), is called juxtacellular nanostimulation. It permits the manipulation the activity of a single cell *in vivo* for a much longer time than what is possible with intracellular stimulation (Houweling et al., 2010; Doron and Brecht, 2015), and it grants a very precise control on the action potentials emitted by the cell (Doose et al., 2016).

After a cell was securely entrained, three types of stimuli were presented at random intervals with different probabilities (fig. 1.5B): i) juxtacellular nanostimulation, ii) a catch trial (no stimulation at all), used to estimate the rat's guessing rate, iii) and microstimulation pulses, used to evaluate whether the rat was attentive and to keep the animal focused on the detection task. If the animal did not respond either to the preceding or to the following microstimulation, nanostimulations and catch trials were excluded from the data analysis. Rats were rewarded by a drop of water if they responded within a response window ranging from 100 ms to 1200 ms after the stimulus onset. The nanostimulation (fig. 1.5C) consisted in a current step of intensity in the nA range and duration 200 ms that generated, depending on the trial, 8 to 16 action potentials (fig. 1.5D).

Figure 1.5E shows in more detail the experimental results obtained from one pyramidal cell in layer 5b, depicted on the left side (dendritic and axonal tree have been reconstructed in red and blue, respectively). In the middle, spikes recorded from the cell for each trial are shown as ticks and the licking responses are indicated as red squares. The rat responded to only one of the eight catch trials (13%), but almost 50% of the times following a single-cell nanostimulation. This particular cell showed one of the strongest increases of the response rate. However, given the relatively low number of trials, the result is not statistically significant on a single-cell basis.

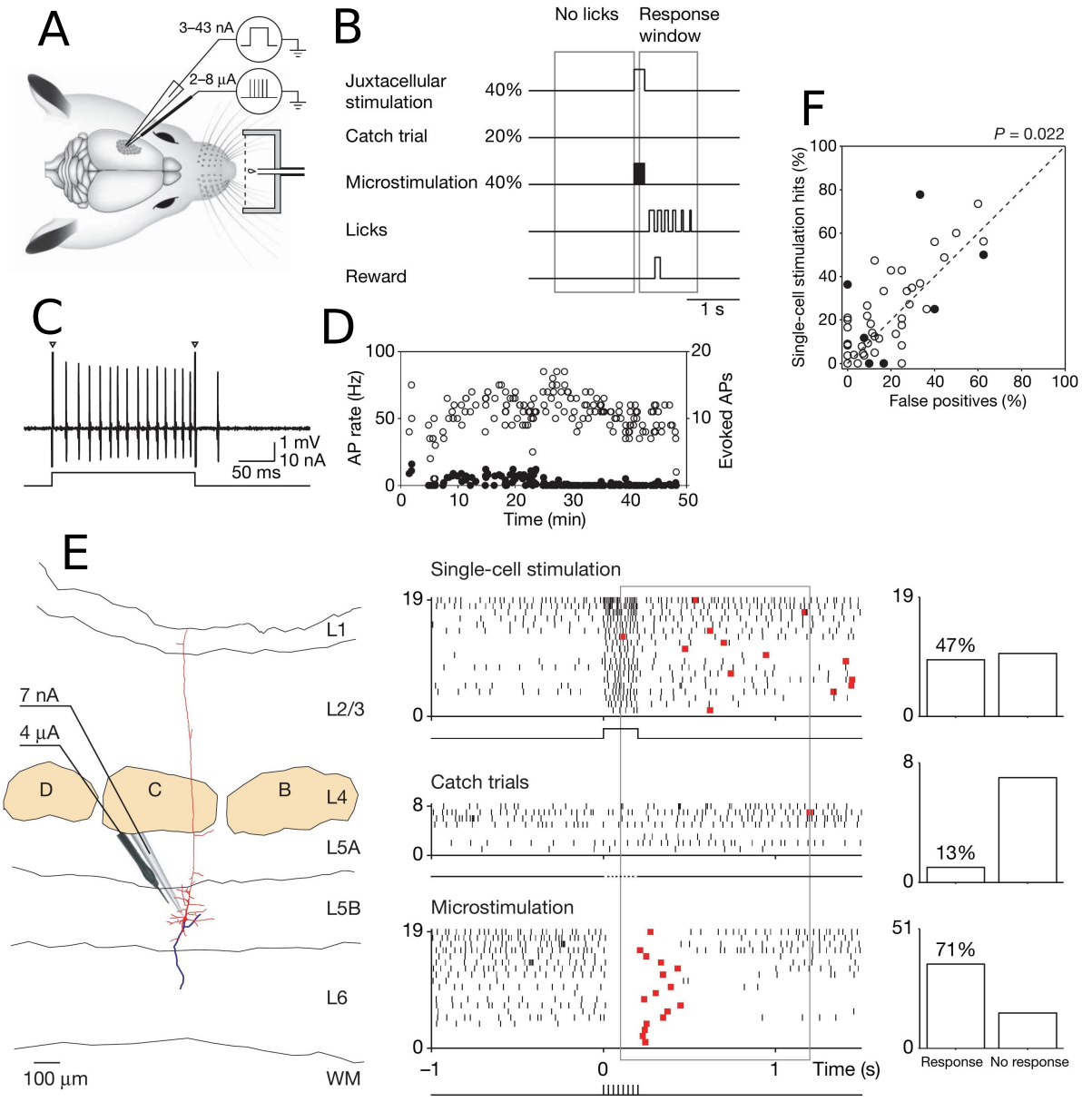

Figure 1.5. – **Experiment by Houweling and Brecht (2008).** **A**: Sketch of experimental setup; rats are awake and can report electrical stimulation in the barrel cortex by licking through the light beam emitted by a light detector (dashed line); correct responses are rewarded by a drop of water. **B**: Three possible stimuli were presented at random times with different probabilities. Licks within the response window were rewarded. **C**: Effect of juxtacellular stimulation. Triangles mark artifacts due to the current jumps. **D**: Spontaneous (black circles) and evoked (open circles) firing rates for a series of nanostimulation trials. **E**: Single-cell stimulation of a single pyramidal neuron from layer 5b, reconstructed on the left (dendritic tree in red, axon in blue). In the middle, spikes are shown as ticks and responses as red squares. **F**: Hit rate vs. false positive rate for all 51 stimulated neurons. Open circles mark putative excitatory neurons, while closed circles mark putative inhibitory neuron. The p-value is obtained from a one-sided paired t-test. Adapted by permission of Springer Nature from Houweling and Brecht (2008).

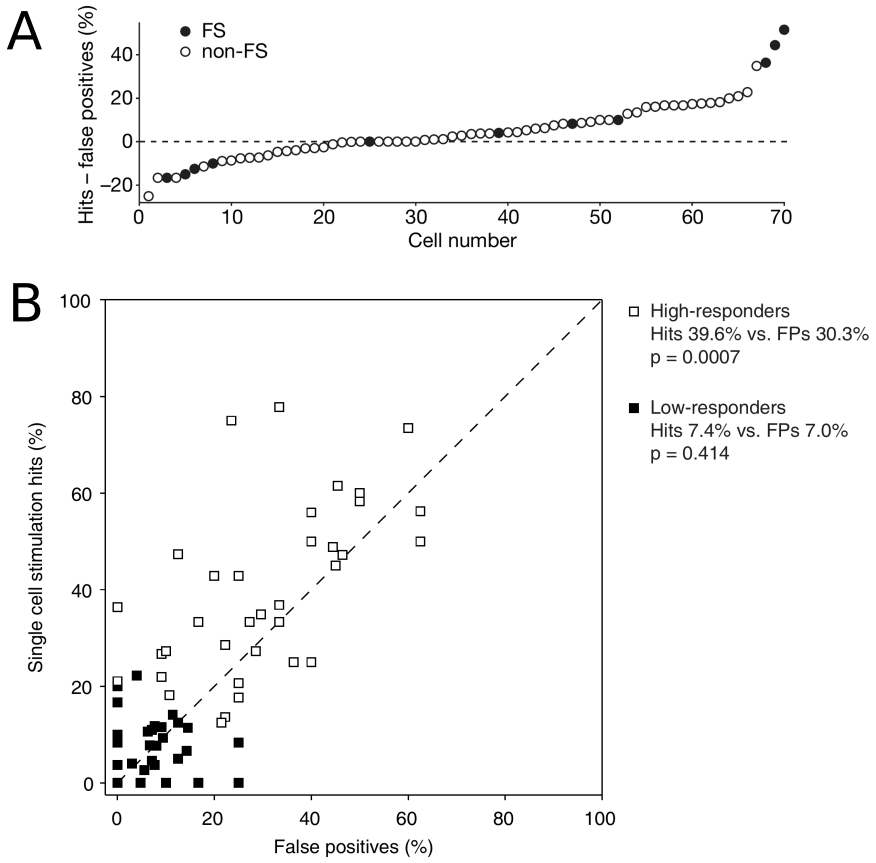

Figure 1.6. – **Putative inhibitory or fast-spiking (FS) cells are more easily detectable than putative excitatory (non-FS) cells. A**: Effect size for all cells. **B**: Hit vs. false positive rate for all cells, divided into two groups, low-responders (overall response rate $\lessapprox 15\%$, closed squares) and high-responders. Adapted from Houweling and Brecht (2008) with permission of Springer Nature.

In total, 51 cells were analyzed and, on average, 30.2 nanostimulation and 17.7 catch trials were included per cell (the number of trials for each cell were rather heterogeneous). The single-cell stimulation hit (correct detection) rate for each cell is plotted as a function of the false positive rate (response rate to catch trials) in fig. 1.5F. Open circles are putative excitatory cells while filled up black dots represent putative inhibitory fast-spiking cells. Averaged over the whole population dataset, the effect size, defined as the difference between hit and false positive rate, was of about 5% ($p = 0.022$, one-sided paired t-test). However, the effect size for fast spiking (FS) neurons, i.e. putative inhibitory cells, was generally stronger than for putative excitatory cells (fig. 1.6A). This result was confirmed by Doron et al. (2014) in a larger dataset.

As a final remark, the effect size was much stronger and statistically significant when animals were not conservative: ordering the cells according to the overall response rate and dividing the dataset in two halves shows that the effect is not significant for low overall response rates

(fig. 1.6B, closed squares) and large and significant for high overall response rates (fig. 1.6B, open squares).

Effect of stimulation parameters on the detectability of single-cell stimulation (Doron et al., 2014)

In a second series of experiments using the same setup, Doron et al. (2014) studied how the detectability of single-cell stimulation was influenced by the parameters of the injected current. They focused on three features of the spike train induced by the nanostimulation: the *total number* of evoked spikes at a given firing rate, the *firing rate* of the stimulated cell for a given number of total spikes, and the *irregularity* of the cell's firing pattern. While firing rate and spike number proved to have hardly any influence on the probability of a behavioral response (only the firing rate had a slightly negative effect), irregular spike trains were found to be significantly easier to detect than regular ones. These results contradict the intuitive expectation that a stronger stimulus should provoke a stronger response, and suggest that the temporal structure of spike trains plays an important role in the way cortical cells encode information. These findings will be presented in more detail in chapter 4, because they are in the focus of the model discussed there.

1.3. Mathematical description of neural activity

This section introduces some definitions useful for the mathematical description of spike trains, i.e. the sequences of action potential emitted by a neuron. These definitions treat spike trains as *stochastic processes* (Stratonovich, 1963; Gardiner, 1985), which is justified by the multiple sources of stochasticity in the neuronal dynamics. Some of them are intrinsic, i.e. due to the biophysical processes underlying the neuronal functioning, such as the random opening and closing of a finite number of ion channels (White et al., 2000). Another example is the stochastic release of neurotransmitter causing the variable amplitude and the transmission failures of synapses (Allen and Stevens, 1994). Another kind of noise is not intrinsic to the neuron, but due to the massive, quasi-stochastic synaptic input to which a cortical neuron is usually subject. The typical cortical neuron has input synapses on the order of thousands. The precise arrival time of the large number of input spikes results from the activity of possibly distant brain regions and from the combination of multiple unknown factors. Therefore, it is usually unpredictable and can be described as a stochastic process, the so-called *synaptic noise*, which is the major noise source in most cortical neurons (Destexhe and Rudolph-Lilith, 2012).

1.3.1. Spike trains, averaging ensembles, first-order statistics

A convenient mathematical representation of the spike train emitted by a neuron is a sum of Dirac delta functions centered on the spike times (Rieke et al., 1999; Gerstner et al., 2014). Hence, if t_i are the spike times, the spike train is

$$x(t) = \sum_i \delta(t - t_i). \tag{1.1}$$

The trial-average of a spike train defines its time-dependent *firing rate*:

$$r(t) = \langle x(t) \rangle , \tag{1.2}$$

where the angular brackets indicate average over different trials. In this thesis, trial-average will imply, depending on the context, averaging over realizations of the synaptic noise, over random initial conditions of the network (which, in some cases, can be considered as an internally generated network input noise), and over different realizations of a frozen noise, e.g. a random realization of the network connectivity. The meaning of angular brackets will be specified case-by-case and indicated by indexes, in case of possible ambiguity.

The time-dependent firing rate of a neuron is the probability density that, in the considered trial, a neuron fires at the time t. In other words, $r(t)\Delta t$ is the probability that a spike is emitted in a short time window centered around t (Rieke et al., 1999). If a set of neurons α of size N_α is considered, its average firing rate is defined as

$$r_\alpha(t) = \left\langle \frac{1}{N_\alpha} \sum_{x \in \alpha} x(t) \right\rangle , \tag{1.3}$$

which denotes the average probability density that one neuron within the network or subnetwork α emits a spike at the time t.

An alternative way of describing a spike train is through the sequence of its interspike intervals, i.e. the time intervals between adjacent spike times $I_i = t_{i+1} - t_i$. The standard measure of the regularity of a spike train is the *coefficient of variation* of its interspike intervals, defined as the standard deviation divided by the mean:

$$\mathrm{CV} = \frac{\sqrt{\langle I^2 \rangle - \langle I \rangle^2}}{\langle I \rangle}. \tag{1.4}$$

The CV is zero for a perfectly regular spike train and it is equal to one for a Poisson process (defined later in this section).

1.3.2. Second-order statistics of spike trains, Fourier transform, spectral measures

Second-order statistics of spike-trains involve two time points. The spike-train *autocorrelation* function is defined as

$$C_{xx}(t,\tau) = \langle x(t+\tau)x(t)\rangle - \langle x(t+\tau)\rangle \langle x(t)\rangle , \tag{1.5}$$

where the first term represents the joint probability density to observe a spike both at time t *and* at $t+\tau$, and the second one is the probability of the same event occurring by chance when independent processes with the same firing rate are considered. In the *stationary* situation, statistics do not depend on the absolute time, but only on time differences, In this case, the firing rate cannot depend on time and the autocorrelation function is a function of one argument, the time lag between two spikes τ:

$$C_{xx}(\tau) = \langle x(t+\tau)x(t)\rangle - r_{\mathrm{sp}}^2, \tag{1.6}$$

where the time t is arbitrary because of the stationarity and $r_{\mathrm{sp}} = \langle x(t)\rangle$ is the stationary firing rate. If the first spike at t is taken as reference and $m(\tau)$ denotes the spike-triggered rate, i.e. the probability density for a spike at $t+\tau$ *given* a spike at t (different from the reference spike, i.e. $m(0) = 0$), then eq. (1.6) can be rewritten as

$$C_{xx}(\tau) = r_{\mathrm{sp}}[\delta(\tau) + m(\tau)] - r_{\mathrm{sp}}^2, \tag{1.7}$$

where the delta function is due to the reference spike.

The degree of correlation between spikes emitted by two different neurons, $x_1(t)$ and $x_2(t)$, is expressed by the spike train *cross-correlation* function, which is defined as (considering again the stationary case):

$$C_{x_1x_2}(\tau) = \langle x_1(t+\tau)x_2(t)\rangle - r_{\mathrm{sp},1}r_{\mathrm{sp},2}, \tag{1.8}$$

where $r_{\mathrm{sp},1}$ and $r_{\mathrm{sp},2}$ are the stationary firing rates of neuron one and two, respectively.

It is often convenient to consider two-point correlations in the Fourier representation. The definition of Fourier transform used in this thesis is

$$\tilde{x}(f) = \lim_{T\to\infty} \tilde{x}_T(f) = \lim_{T\to\infty} \int_0^T \mathrm{d}t\, e^{2\pi i f t} x(t). \tag{1.9}$$

In the stationary case, it can be shown that $\langle \tilde{x}(f \neq 0)\rangle = 0$ and that (the asterisk denotes the complex conjugate)

$$\langle \tilde{x}(f)\tilde{x}^*(f')\rangle = \delta(f-f')S_{xx}(f), \tag{1.10}$$

i.e. that the autocorrelation of the Fourier transformed spike train (and of any stationary stochastic process) is zero when two different frequencies f and f' are considered and that it diverges as $T \to \infty$ when $f = f'$. The proportionality factor $S_{xx}(f)$ is the spike-train *power-spectrum*

$$S_{xx}(f) = \lim_{T\to\infty} S_{xx,T}(f) = \lim_{T\to\infty} \frac{\langle \tilde{x}_T(f)\tilde{x}_T^*(f)\rangle}{T}. \tag{1.11}$$

Analogously, the cross-spectrum between the spike trains generated by two neurons $x_1(t)$ and $x_2(t)$ is defined as

$$S_{x_1x_2}(f) = \lim_{T\to\infty} S_{x_1x_2,T}(f) = \lim_{T\to\infty} \frac{\langle \tilde{x}_{1,T}(f)\tilde{x}_{2,T}^*(f)\rangle}{T}. \tag{1.12}$$

Note that the cross-spectrum is conjugate symmetric, i.e. swapping the role of the two neurons in eq. (1.12) is equivalent to taking the complex conjugate of the spectrum.

In the next chapters, the average spike-train power- and cross-spectra within a network or a specific subset of a network are often considered. In this case, angular brackets will indicate averaging both over trials and over different neurons or neuron pairs within the considered subset. The duration of the spike train (indicated by T) is never infinite in practice so that, strictly speaking, all spectral measures computed from simulations depend on T. However, if T is large enough, it will be assumed that the dependence on T is weak and that it can be omitted from the notation.

Power-spectrum and autocorrelation function are a Fourier transform pair, a fact that is known as Wiener-Kirchin theorem (Gardiner, 1985):

$$S_{xx}(f) = \int_{-\infty}^{+\infty} \mathrm{d}\tau\, e^{2\pi i f\tau} C_{xx}(\tau) \tag{1.13}$$

$$C_{xx}(\tau) = \int_{-\infty}^{+\infty} \mathrm{d}f\, e^{-2\pi i f\tau} S_{xx}(f) \tag{1.14}$$

One corollary of eq. (1.14) is that the variance of any stationary process can be expressed as the integral over its power spectrum:

$$\sigma_x^2 = \left\langle x^2(t)\right\rangle - \left\langle x(t)\right\rangle^2 = C_{xx}(0) = \int_{-\infty}^{+\infty} \mathrm{d}f\, S_{xx}(f) e^{-2\pi i f(t=0)} = \int_{-\infty}^{+\infty} \mathrm{d}f\, S_{xx}(f). \tag{1.15}$$

Several properties of a stochastic process can be read out from its power spectrum. For instance, an oscillation appears as a peak in the power spectrum. The peak will be sharp for a coherent oscillation (a pure sine wave produces one pair of delta functions in the power spectrum), and broad for noisier oscillations. In the case of a spike train, the high- and low-

frequency limits of the power spectrum give information about the firing rate and regularity. For non-pathological cases (see Droste and Lindner, 2017b, for one interesting counterexample), it can be shown that the high-frequency limit of the power spectrum of a spike train is the firing rate (Lindner, 2013)

$$\lim_{f\to+\infty} S_{xx}(f) = r_{\mathrm{sp}}. \tag{1.16}$$

The low-frequency limit of the power spectrum is linked to the regularity and to the correlations between inter-spike intervals (Cox and Lewis, 1966):

$$\lim_{f\to 0} S_{xx}(f) = r_{\mathrm{sp}}\,\mathrm{CV}^2\left(1+2\sum_k \rho_k\right) \tag{1.17}$$

where ρ_k is the serial correlation coefficient at lag k, i.e. the linear correlation coefficient between interspike intervals separated by $k-1$ intervals.

1.3.3. Poisson neuron *alias* shot noise

The simplest way to model a spiking neuron is to ignore the internal dynamics completely and to generate spikes at independent, random times. In practice, a very short time discretization Δt needs to be chosen, and for each time step a uniform random number η in the interval $(0,1)$ is drawn, independently of the past. If $\eta < r(t)\Delta t$, a spike is generated. If the probability of firing per unit time is constant in time ($r(t) = r_{\mathrm{sp}}$), the resulting spike train is a realization of a Poisson process. The number of spikes n occurring in any (not necessarily short) time interval of length Δt is distributed according to the Poisson distribution:

$$P(n,\Delta t) = e^{-r_{\mathrm{sp}}\Delta t}\frac{(r_{\mathrm{sp}}\Delta t)^n}{n!}. \tag{1.18}$$

When $\Delta t \ll 1/r_{\mathrm{sp}}$, the probability of observing one spike in Δt is $P(1,\Delta t) \approx r_{\mathrm{sp}}\Delta t$, and the probability of observing more than one spike is negligible, so that $P(0,\Delta t) \approx 1 - r_{\mathrm{sp}}\Delta t$, consistent with the generation algorithm described above. Because all spike times are independent of each other, the spike-triggered rate of a Poisson process is equal to the stationary firing rate, i.e. $m(\tau) = r_{\mathrm{sp}}$. From eq. (1.7) it follows that the autocorrelation of a Poisson spike train is

$$C_{xx}(\tau) = r_{\mathrm{sp}}\delta(\tau). \tag{1.19}$$

The Fourier transform of eq. (1.19) is

$$S_{xx}(f) = r_{\mathrm{sp}}, \tag{1.20}$$

i.e. the power spectrum of a Poisson process is flat. By loose analogy with the spectrum of white light, the term *white* is used to describe a stochastic process with a flat power spectrum, or, equivalently, with temporal autocorrelation as in eq. (1.19).

In this thesis, Poisson spike trains will be used to mimic synaptic noise, i.e. the spiking background from other brain areas. This kind of input noise is also termed *shot noise* to emphasize that it affects the system by delivering finite "kicks".

1.4. Neuron models

As briefly outlined in section 1.1, neurons are morphologically complex structures, whose function relies on a large number of stochastic ion channels of multiple types. State-of-the-art biologically detailed computational models of single neurons can feature thousands of differential equations and tens of thousands of parameters (Almog and Korngreen, 2016). On the other end of the structural complexity spectrum, there are *point neuron models*.

Point neuron models, integrate-and-fire models

In point neuron models, the spatial structure of neurons is neglected and the membrane potential inside the soma is the main (if not the only) state variable of interest. If the total capacitance of the neuronal membrane is C_m and the membrane potential inside the soma is indicated by v, the current-balance equation is

$$C_m \frac{\mathrm{d}v}{\mathrm{d}t} = -I_m + I_{\mathrm{syn}} + I_0, \tag{1.21}$$

where on the left hand side the capacitive current has been isolated and all other terms have been brought to the right hand side. The term I_{syn} represents the total synaptic currents entering the soma either directly or from the dendritic tree, i.e. the input from other neurons. The term I_0 represents an externally applied current. The term I_m describes the total current flowing through ion channels within the membrane. Its negative sign is a convention, and it contains also the currents responsible for spike generation. In the famous model by Hodgkin and Huxley (1952), these currents are the combination of one sodium and one potassium current governed by the combined activation and inactivation of three voltage-dependent conductances. This model is a four-dimensional, non-linear system of equations, which makes it difficult to study analytically. Furthermore, its numerical integration requires a fine time step to ensure that the interplay of the currents generates an action potential.

In *integrate-and-fire* models (Dayan and Abbott, 2001; Gerstner et al., 2014), the spike-generating current is replaced by a phenomenological function $f(v)$ complemented with a fire-and-reset rule: whenever the voltage reaches a threshold value, a spike is emitted and v is reset

to a prescribed value:

$$C_m \frac{\mathrm{d}v}{\mathrm{d}t} = f(v) + I_{\mathrm{syn}} + I_0. \tag{1.22}$$

Typical choices for the function $f(v)$ are either a constant, a linear, a quadratic, or an exponential function, which model the subthreshold dynamics and, in the case of the quadratic or exponential integrate-and-fire model, the upstroke of the action potential.

1.4.1. The leaky integrate-and-fire neuron

In this thesis, the leaky integrate-and-fire (LIF) model will be used, in which $f(v)$ is a linear function representing the leak current, which tends to bring the membrane potential back to the resting potential. Therefore, in the LIF model, eq. (1.22) takes the form

$$C_m \frac{\mathrm{d}v}{\mathrm{d}t} = -g_L(v - E_L) + I_{\mathrm{syn}} + I_0, \tag{1.23}$$

where g_L is the leak conductance of the membrane and E_L is the resting potential of the neuron. If voltages are measured with respect to the resting potential, then $E_L = 0$, which disappears from the equation. Furthermore, by introducing the membrane resistance $R_m = 1/g_L$ and the *membrane time constant* $\tau_m = C_m/g_L = R_m C_m$, eq. (1.23) can be rewritten as

$$\tau_m \frac{\mathrm{d}v}{\mathrm{d}t} = -v + R_m I_{\mathrm{syn}} + R_m I_0. \tag{1.24}$$

In the LIF model, the action potential upstroke is not modeled explicitly, and a spike is emitted whenever a hard threshold v_T is reached. After emitting one spike, real neurons enter an *absolute refractory period*, in which they cannot fire again, no matter how strong the input current is (Kandel et al., 2000). To mimic absolute refractoriness in the LIF, the voltage is clamped to the reset voltage v_R for the duration of the refractory period τ_{ref} and then let free to evolve. Figure 1.7A shows an example voltage trace produced by a LIF neuron, for which the constant external input (indicated by the dotted line) is above the firing threshold (marked by a red dashed line) and the synaptic input is absent $R_m I_{\mathrm{syn}} = 0$. Except for the first interval, which depends on the initial conditions, the model produces a perfectly regular spike train. The spike peak was added only for illustration and it does not result from the model dynamics.

Current-based synapses

The term I_{syn} in eqs. (1.21) to (1.24) represents the total synaptic current entering the soma caused by the arrival of spikes at the neuron's input synapses. One of the most minimalistic ways to model the effect of presynaptic spikes on the neuron is by instantaneous jumps of the voltage by a quantity representing the amplitude of the postsynaptic potential.

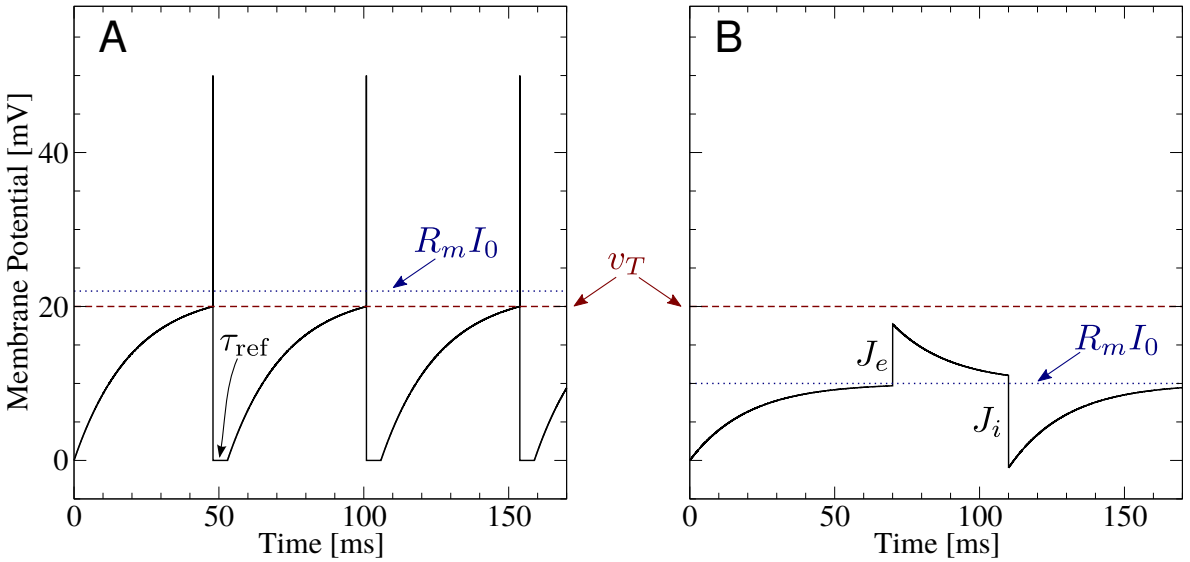

Figure 1.7. – **Example voltage traces generated by a leaky integrate-and-fire neuron and effect of instantaneous current-based synapses**. **A**: Constant mean input ($R_m I_0 = 22\,\text{mV}$, dotted line) drives the voltage towards the threshold $v_T = 20\,\text{mV}$ (dashed line). When $v(t)$ reaches the threshold, it is reset to v_R and a spike is emitted. The voltage is clamped to the reset point for the duration of the refractory period τ_{ref} before it can evolve further. **B**: The constant mean input ($R_m I_0 = 10\,\text{mV}$, dotted line) is below the firing threshold and this neuron could only reach the threshold by effect of its synaptic input. One excitatory and one inhibitory input spikes make the voltage jump by J_e and $-J_i$, respectively. Other parameters: $\tau_m = 20\,\text{ms}$, $\tau_{\text{ref}} = 5\,\text{ms}$, $v_T = 20\,\text{mV}$, $v_R = 0\,\text{mV}$.

Suppose that one presynaptic excitatory neuron spikes at $t = \hat{t}_e$, that one presynaptic inhibitory neuron spikes at $t = \hat{t}_i$, and that the corresponding input synapses cause a postsynaptic potential of amplitude J_e and J_i, respectively. In this thesis, non-linear interactions in dendrites are neglected, so that the total effect of input spikes can be represented by the sum of the contribution from each synapse. According to this model, the evolution equation for the LIF neuron receiving these two spikes reads

$$\tau_m \frac{\mathrm{d}v}{\mathrm{d}t} = -v + \tau_m J_e \delta(t - \hat{t}_e - D) - \tau_m J_i \delta(t - \hat{t}_i - D) + R_m I_0, \tag{1.25}$$

where the factor τ_m multiplying the Dirac delta function ensures that v jumps by J_e (-J_i) at $t = \hat{t}_e + D$ ($t = \hat{t}_i + D$), and D represents a transmission delay due to multiple factors, such as the propagation along the presynaptic axon, the time for neurotransmitter release and diffusion, and the dendritic propagation. The effect of the two spikes is shown in fig. 1.7B, in which the constant input term (dotted line) is below the firing threshold (red dashed line).

Leaky integrate-and-fire neuron driven by shot noise

Consider now a LIF neuron receiving a barrage of input spikes. If t_ℓ and t_k indicate the arrival times of excitatory and inhibitory spikes, respectively, the time evolution of the neuron reads (in the rest of this section the fire-and-reset rule is implicitly assumed for any differential equation for v)

$$\tau_m \frac{\mathrm{d}v}{\mathrm{d}t} = -v + \tau_m \sum_\ell a_{e,\ell}\delta(t - t_\ell) - \tau_m \sum_j a_{i,k}\delta(t - t_k) + R_m I_0, \tag{1.26}$$

where $a_{e,\ell}$ ($a_{i,k}$) indicates the amplitude of the ℓth (kth) excitatory (inhibitory) spike. Richardson and Swarbrick (2010) derived several analytical expressions for the case that the input spike trains are a Poisson process and the amplitudes are exponentially distributed. The analytical result that will be applied most often in this thesis is the stationary firing rate:

$$\phi_{sn}(a_e, a_i, R_i, R_e, I_0) = \left(\tau_{\mathrm{ref}} + \tau_m \int_0^{1/a_e} \frac{\mathrm{d}s}{s} Z_0^{-1}(s) \left[\frac{e^{s(v_T - R_m I_0)}}{1 - a_e s} - e^{s(v_R - R_m I_0)} \right] \right)^{-1}, \tag{1.27}$$

where a_e (a_i) indicates the mean of the exponential distribution from which excitatory (inhibitory) amplitudes are drawn, R_e (R_i) indicates the firing rate of the excitatory (inhibitory) Poissonian input spike train. Furthermore, $Z_0^{-1}(s) = (1 - a_e s)^{\tau_m R_e}(1 + a_i s)^{\tau_m R_i}$.

Leaky integrate-and-fire neuron driven by Gaussian white noise

The classical approach to tackle eq. (1.26) is to model the summed effect of all input spikes as a Gaussian white noise, which is known as the *diffusion approximation* (Ricciardi and Sacerdote, 1979; Lánskỳ and Lanska, 1987). For this approximation to work well, two conditions must be satisfied: i) the input rate is large enough that many input spikes arrive, on average, in a short (compared to the membrane time constant) time interval; ii) the spike amplitudes are not too large. As a consequence of the diffusion approximation, the LIF model becomes mathematically equivalent to a Langevin equation (Gardiner, 1985) with a fire-and-reset rule:

$$\tau_m \frac{\mathrm{d}v}{\mathrm{d}t} = -v + \mu + \sqrt{2D}\xi(t), \tag{1.28}$$

where μ is a constant parameter describing the mean input, D is the parameter setting the noise intensity and $\xi(t)$ is a Gaussian white noise process with zero mean $\langle \xi \rangle = 0$ and unit intensity, i.e. the correlation function of the noise is $\langle \xi(t')\xi(t) \rangle = \delta(t - t')$. To approximate the dynamics in eq. (1.26), the value of μ and D must be chosen as follows (Richardson and Swarbrick, 2010):

$$\mu = \tau_m(a_e R_e - a_i R_i) + R_m I_0, \tag{1.29}$$

$$D = R_e \tau_m^2 a_e^2 + R_i \tau_m^2 a_i^2. \tag{1.30}$$

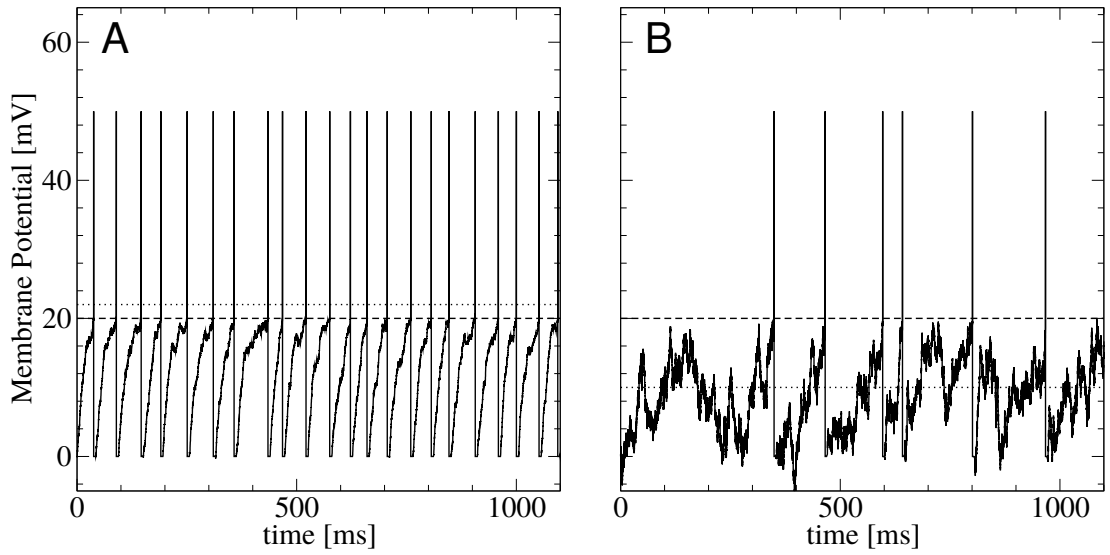

Figure 1.8. – **Firing regimes of a leaky integrate-and-fire neuron driven by Gaussian white noise**. **A**: Mean-driven regime ($\mu = 22\,\mathrm{mV}$, $D = 0.1\,\mathrm{mV^2ms}$); **B**: noise-driven regime ($\mu = 10\,\mathrm{mV}$, $D = 1\,\mathrm{mV^2ms}$). In both panels, the values of μ is marked by a dotted line and the firing threshold is shown as a dashed line. Other parameters: $\tau_m = 20\,\mathrm{ms}$, $v_T = 20\,\mathrm{mV}$, $v_R = 0\,\mathrm{mV}$, $\tau_{\mathrm{ref}} = 5\,\mathrm{ms}$.

The noise intensity differs by a factor 2 from the case of fixed input amplitudes (Lindner, 2013; Droste, 2015).

Depending on the value of μ with respect to the threshold v_T, two main firing regimes can be distinguished. When $\mu > v_T$ the neuron is in the *mean-driven regime*, shown in fig. 1.8A. In this regime, the output spike train is typically rather regular (except for extreme values of the noise) and the neuron can fire at any noise level (and even in the absence of noise, as in fig. 1.7A). When $\mu < v_T$, only noise fluctuations can bring the voltage above the threshold. As a consequence, this *fluctuation-driven* regime is characterized by lower firing rates and irregular inter-spike intervals, as it can be seen in fig. 1.8B. These two firing regimes are not restricted to the case of Gaussian white noise, but are analogous in the case of shot-noise input, where the role of μ is played by the total mean input, calculated as in eq. (1.29).

The leaky integrate-and-fire neuron driven by Gaussian white noise has been studied for several decades and several properties of the spontaneous activity and of the firing-rate response to stimulation have been calculated (Brunel et al., 2001; Lindner and Schimansky-Geier, 2001; Lindner, 2002; Voronenko and Lindner, 2017). In this thesis, only the formula for the stationary firing rate of the neuron will be used for a comparison to the shot-noise theory by Richardson and Swarbrick (2010). The firing rate of the LIF driven by Gaussian white noise reads (Ricciardi

and Sacerdote, 1979)

$$\phi_{wn}\Big(\mu, D\Big) = \left(\tau_{\rm ref} + \tau_m \int\limits_{\frac{\mu - v_T}{2D}}^{\frac{\mu - v_R}{2D}} {\rm d}z \, e^{z^2} \, {\rm erfc}(z) \right)^{-1} , \tag{1.31}$$

where erfc(z) is the complementary error function.

1.4.2. Input-output *(f-I)* curves, linear response to signals, DC susceptibility

Consider the situation described by eq. (1.26), i.e. a LIF neuron with spiking input background and a constant input term $R_m I_0$. The stationary firing rate of the neuron as a function of the constant input is sometimes indicated as *f-I curve* of the neuron. In the case of Poisson input with exponentially distributed amplitudes, the f-I curve for the LIF model is given by eq. (1.27), regarded as a function of $R_m I_0$.

Some examples of f-I curves are shown in fig. 1.9, in which the firing rate of the neuron is shown as a function not of $R_m I_0$ itself, but of $R_m \Delta I_0 = R_m I_0 - R_m \hat{I}_0$, which is the shift with respect to the "operating point" of the neuron, i.e. to some reference value $R_m \hat{I}_0$. The three solid lines of different colors in fig. 1.9A refer to three different input noise levels, in which amplitude and rate of the shot-noise background is varied. The parameter values are chosen to mimic the recurrent input received by a neuron embedded in the network model considered in section 2.4. For each case, the numerical value of the mean excitatory amplitude is indicated in the inset as J, while the mean inhibitory amplitude is proportional to it. Both excitatory and inhibitory input rates are proportional to $r_{\rm in}$, which represents the average network firing rate and is also indicated in the inset (further details on all parameter values can be found in the figure caption). When the input amplitudes are weak and the input rate is high (black solid line), the f-I is rather nonlinear and well approximated by the diffusion approximation (black dotted line). When the synaptic input weight is raised and the input rate is reduced, the overall noise intensity increases and the f-I becomes more linear (red solid line). For this case, discrepancies between the shot noise theory and the diffusion approximation (red dotted line) can be seen in the intermediate range. If the shot-noise amplitude is further increased and the input rate further reduced, the f-I curve becomes rather flat (blue solid line) and the diffusion approximation (blue dotted line) overestimates the firing rate everywhere.

Although in fig. 1.9A the f-I curve for the case of largest noise (blue solid line) looks rather linear, if the range is expanded (fig. 1.9B) it becomes apparent that the range shown in fig. 1.9A is where the curvature is largest (excluding the saturation at $1/\tau_{\rm ref}$ for extremely large inputs, not shown). The blue line in fig. 1.9B the same as the blue solid line in fig. 1.9A, but plotted over a wider range. In chapter 2, a network with additional excitatory external input is considered,

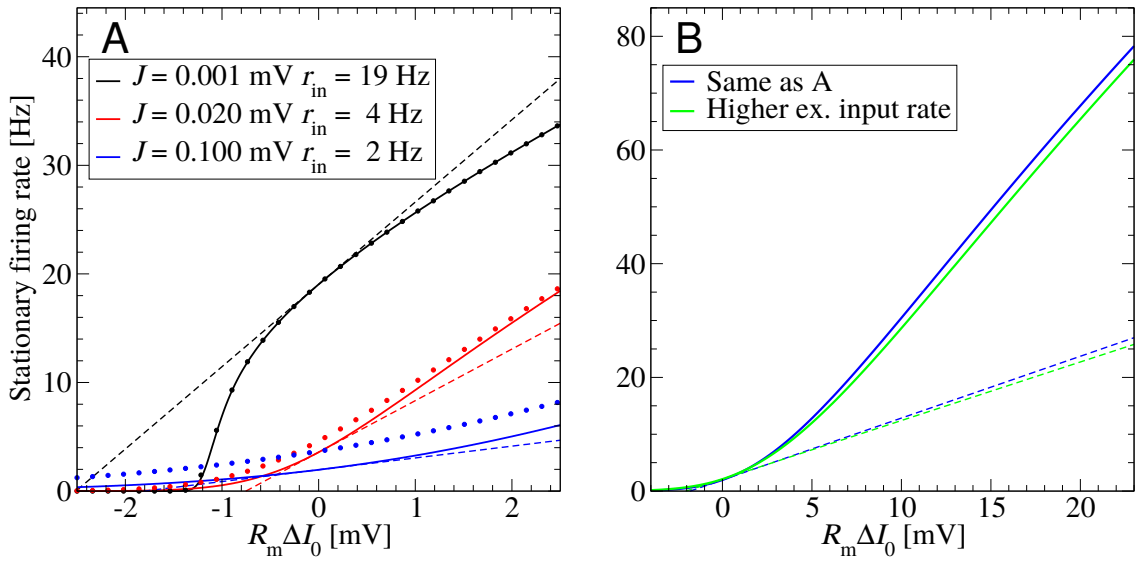

Figure 1.9. – **Noise can linearize the f-I curve and reduce the DC susceptibility of a leaky integrate-and-fire neuron with white noise input background.** Stationary firing rate of LIF neuron as a function of a shift from a reference value, i.e. $R_m\Delta I_0 = R_m I_0 - R_m\hat{I}_0$. Solid lines refer to shot-noise with exponentially distributed kicks and are computed by using eq. (1.27). Dotted lines represent the diffusion approximation and are computed by combining eqs. (1.29) to (1.31). Dashed lines are the tangents to the solid line at the point $\Delta I_0 = 0$, and are computed by using the derivative of eq. (1.27). The explicit expression is eq. (A.6) on p. 199. The cases shown here correspond to injecting an external current into a neuron receiving shot-noise background of the same amplitude and total input rate as a neuron in the network of section 2.4. In panel **A** it can be seen how the amplitude of the input noise influences the DC susceptibility. In panel **B**, the x-axis terminates at the value of current injection used to mimic single-cell stimulation (see section 2.2) and the two f-I curves change slightly when the excitatory input rate is increased to imitate the two cases considered in section 2.4 (autonomous network and with input from the thalamus). Parameters (**A**): $\tau_m = 20\,\text{ms}$, $\tau_{\text{ref}} = 2\,\text{ms}$, $R_m\hat{I}_0 = 22\,\text{mV}$, $a_e = J$, $a_i = gJ = 7\,J$, $R_e = C_E r_{\text{in}} = 4000\,r_{\text{in}}$, $R_i = C_I r_{\text{in}} = 1000\,r_{\text{in}}$. Parameters (**B**): blue line as in **A**; green line: $R_m\hat{I}_0 = 5.2\,\text{mV}$, $R_e = C_{\text{ext}} r_{\text{ext}} + C_E r_{\text{in}} = 700\cdot 12 + 4000\cdot 2\text{Hz}$, otherwise as for the blue line in **A**.

which leads to a doubled excitatory input noise (while keeping the mean input constant). This case is plotted with the green solid line, which shows that the f-I does not change much in this case. All dashed lines in fig. 1.9 are the tangents to the f-I curve for $R_m \Delta I_0 = 0$, and represent the linear response to slow inputs, as explained below (the following derivation was taken from Lindner, 2013).

Suppose that the neuron receives an additive signal input signal, which can be represented by making the constant input term in eq. (1.26) time dependent, i.e. $R_m I_0(t) = \mu(t) = \mu_0 + s(t)$, where the constant term μ_0 summarizes the baseline input and $\langle s(t) \rangle = 0$. The signal can be either deterministic or a realization of a stochastic process. In the latter case, the angular brackets appearing in the following equations indicate averaging over the noise ensemble, i.e. a particular frozen realization of the signal is considered. If the signal is weak, the output firing rate of the neuron can be described by the *linear-response* ansatz:

$$\langle x(t) \rangle = r(t) = r_{\mathrm{sp}} + \int_{-\infty}^{+\infty} \mathrm{d}t' \, K(t') s(t - t'), \tag{1.32}$$

where $r_{\mathrm{sp}} = \phi_{sn}(\mu_0)$ is the output firing rate in the absence of the signal and $K(t)$ is the linear response kernel, which describes the effect of the weak signal at past times. From inspection of eq. (1.32), it is clear that the linear response kernel must be zero for negative times if causality is to be preserved. In other words, $K(t < 0) = 0$ must be imposed to prevent future times to influence the present value of the signal. The convolution in eq. (1.32) becomes a simple product by applying the Fourier transform:

$$\langle \tilde{x}(f) \rangle = \tilde{r}(f) = r_{\mathrm{sp}} \delta(f) + \chi(f) \tilde{s}(f), \tag{1.33}$$

where

$$\chi(f) = \int_{-\infty}^{+\infty} \mathrm{d}t \, e^{2\pi i f t} K(t), \tag{1.34}$$

is the firing-rate *susceptibility* with respect to an additive signal. Suppose now that the signal varies slowly compared to the internal time scales of the system; in this case, the firing rate of the neuron at each time can be approximated by the stationary firing rate at that time, as in an adiabatic approximation:

$$r(t) \approx \phi_{sn}(\mu_0 + s(t)). \tag{1.35}$$

If the signal appearing in eq. (1.35) is both slow *and* weak, a Taylor expansion truncated to the first order yields

$$r(t) \approx \phi_{sn}(\mu_0 + s(t)) \approx r_{\mathrm{sp}} + \frac{\mathrm{d}\phi_{sn}}{\mathrm{d}\mu} s(t). \tag{1.36}$$

The linear-response ansatz eq. (1.32) already assumes that the signal is weak. If the signal appearing there is also slow compared to the memory of the system, then $s(t-t') \approx s(t)$ for any time difference over which the kernel is substantially different from zero. Therefore, under the assumption of a slow signal, eq. (1.32) becomes

$$r(t) \approx r_{\mathrm{sp}} + \int_{-\infty}^{+\infty} \mathrm{d}t'\, K(t') s(t-t') \approx r_{\mathrm{sp}} + s(t) \int_{-\infty}^{+\infty} \mathrm{d}t'\, K(t') = r_{\mathrm{sp}} + \chi(0) s(t). \tag{1.37}$$

Equations (1.36) and (1.37) were obtained under the same assumption of weak, slow signal, so that they must be equivalent. Comparing them leads to the identity

$$\chi(0) = \frac{\mathrm{d}\phi_{sn}}{\mathrm{d}\mu}. \tag{1.38}$$

The linear response to slow-varying signals $\chi(0)$ is sometimes termed DC susceptibility to underscore its relationship with the f-I curve.

Although the relation between the low-frequency limit of the susceptibility and the derivative of the stationary input-output relationship in eq. (1.38) holds in general, in this thesis, only the DC susceptibility for the LIF neuron with exponentially distributed shot-noise input will be used. Therefore, $\chi(0)$ will be used to indicate this particular susceptibility, as in eq. (1.38). For each example case considered in fig. 1.9, $\chi(0)$ at the operation point $R_m \Delta I_0 = 0$ has been plotted as a dashed line. The three curves in fig. 1.9A show that when the noise intensity is reduced by decreasing the synaptic weights, the susceptibility first increases and then it saturates. In other words, for a weak slow signal of given strength, the modulation of the output firing rate is larger in the case of weaker noise, a fact that will help to interpret some results in chapter 2.

1.5. Networks

A network is a collection of items (*vertices*) connected by links (*edges*) used to describe a system of discrete elements interacting in some way (Newman, 2003). In mathematical terms, the structure of a network is described by a *graph*, which is an ordered pair of sets (V, E), where V is the set of vertices, and $E \subseteq V \times V$ is the set of edges, i.e. pairs of nodes. If these pairs are ordered, the graph is *directed*. A graph can be viewed as the "skeleton" of a network, as it does not carry any information about the items represented by its vertices, but only about the presence and the properties of the connections between them. Networks of neurons are naturally represented by a directed graph, in which each neuron is a vertex and each edge denotes the presence of a synapse. The number of edges terminating at a vertex of a directed graph is called *in-degree* of that vertex, while the number of edges originating from it is termed the *out-degree*.

An alternative way of representing a directed graph instead of listing its vertices and edges is

through a connection matrix J_{ij}, which has a non-zero entry only if there is an edge connecting vertex (neuron) j to vertex i (the established notation, which will be adopted in this thesis, employs the row index for the target and the column index for the source). If the connection matrix only takes the values zero or one, it represents the network's topology and it is termed an *adjacency matrix*. In a *weight matrix*, the element J_{ij} represents the strength of the connection from j to i and can take any value, where zero indicates the absence of a connection.

1.5.1. Regular random graph

The topology of most networks considered in this thesis is a *regular random graph*, which has two fundamental properties: all vertices have the same in-degree (*regular graph*), and the origin of each edge is chosen at random independently of the others. To construct such a network, one can proceed as follows. Let C be the prescribed in-degree and N the number of vertices (neurons). Each neuron of the network is considered in turn as target neuron. Edges terminating at that neuron are added by selecting at random source neurons independently and with equal probability, until the number of input connections is C. Both self-connections (called *autapses* in neuroscience) and multiple connections between the same (ordered) pair of neurons were excluded from all network models considered in this thesis. Multiple synaptic contacts between the same pair of neurons are commonplace in the rat somatosensory cortex (Schnepel et al., 2014; Schoonover et al., 2014; Koelbl et al., 2015) and autapses have also been observed (Lübke et al., 1996). However, it will be assumed that a single connection in the model can represent the overall effect of all synapses connecting two neurons, and that autapses have no major impact on the network dynamics. By construction, the in-degree is fixed, while the out-degree of each neuron follows the binomial distribution for $N-1$ draws with success probability $p_c = C/(N-1)$.[1] It follows that the mean out-degree is C.

1.5.2. Random network of leaky integrate-and-fire neurons (Amit-Brunel Network)

A set of excitatory and inhibitory leaky integrate-and-fire neurons connected by current-based synapses (see p. 23) according to a regular random graph is a classical spiking network model (Amit and Brunel, 1997a,b), which is an enormous simplification of a cortical network. Still, the Amit-Brunel model can produce a vast repertoire of dynamical phenomena, including synchronized and desynchronized firing regimes, possibly coexisting with collective oscillations (Brunel, 2000).

In its most essential version, the Amit-Brunel network consists of N_E excitatory and $N_I =$

[1] A connection between two randomly chosen neurons is present with probability $p_c = C/(N-1)$, because the C inputs to each neuron are chosen randomly from $N-1$ possibilities. Because any neuron has $N-1$ potential targets and a connection is present with probability p_c, the total number of outputs follows a binomial distribution for $N-1$ draws with probability p_c.

γN_E LIF neurons. Each neuron receives input from C_E excitatory and $C_I = \gamma C_E$ inhibitory neurons, chosen at random. All excitatory weights are equal to J, and all inhibitory connections have strength gJ. In other words, all rows of the weight matrix have exactly C_E non-zero elements equal to J and C_I non-zero elements equal to $-gJ$. Furthermore, all non-zero elements of each column of the weight matrix J_{kj} are either all J or all $-gJ$, i.e. each neuron can be only excitatory or inhibitory (Dale's principle). In addition to the input from within the network, each neuron receives external input shot noise with fixed amplitude J. Hence, the kth neuron evolves according to

$$\tau_m \frac{\mathrm{d}v_k}{\mathrm{d}t} = -v_k + \tau_m \sum_j^C J_{kj} x_j(t-D) + \tau_m \sum_\ell^{C_{\mathrm{ext}}} J x_\ell^{\mathrm{ext}}(t) \tag{1.39}$$

with the fire-and-reset rule. In eq. (1.39), $x_j(t)$ is the spike train fired by the jth neuron, $x_\ell^{\mathrm{ext}}(t)$ are C_{ext} Poissonian spike trains with rate r_{ext}, and D is the transmission delay.

When the number of inputs per neuron is large and the weights are small compared to the distance from the reset to the threshold voltage, the diffusion approximation can be employed to describe the input to each neuron. In this way, eq. (1.39) is replaced by

$$\tau_m \frac{\mathrm{d}v_k}{\mathrm{d}t} = -v_k + \mu_{\mathrm{eff}} + \sqrt{2D_{\mathrm{eff}}}\xi(t), \tag{1.40}$$

where $\xi(t)$ is a Gaussian white noise process with zero mean and unit intensity (see also p. 25). The effective mean input and the effective noise intensity are (Amit and Brunel, 1997a)

$$\mu_{\mathrm{eff}} = \tau_m J C_E (1 - g\gamma) r_{\mathrm{sp}} + \tau_m J C_{\mathrm{ext}} r_{\mathrm{ext}}, \tag{1.41}$$

$$2D_{\mathrm{eff}} = \tau_m^2 J^2 C_E (1 + g^2\gamma) r_{\mathrm{sp}} + \tau_m^2 J^2 C_{\mathrm{ext}} r_{\mathrm{ext}}. \tag{1.42}$$

and depend on r_{sp}, the spontaneous firing rate of the network. The output firing rate of each neuron can be expressed as a function of μ_{eff} and D_{eff}:

$$r_{\mathrm{sp}} = \phi_{wn}\Big(\mu_{\mathrm{eff}}, D_{\mathrm{eff}}\Big), \tag{1.43}$$

where $\phi_{wn}\Big(\mu_{\mathrm{eff}}, D_{\mathrm{eff}}\Big)$ is the firing rate of a LIF neuron driven by Gaussian white noise, eq. (1.31). Self-consistency of input and output firing rates allows to find the spontaneous firing rate r_{sp}, by solving the system consisting of eqs. (1.41) to (1.43) numerically.

Adding a perturbation to eqs. (1.41) and (1.42) and to the linearization of eq. (1.43) yields a condition for the linear stability of the fixed point (Amit and Brunel, 1997a; Ledoux and Brunel, 2011). Brunel (2000) conducted a full analytical study of the transitions of the system to synchronized states by means of a perturbational analysis of the solution to the Fokker-Planck

equation associated to eq. (1.40). Note that the self-consistency imposed by these approaches is limited to the firing rate of the network, as the input to each neuron is approximated as white noise, although the output spike train of any neuron in the network is not. Numerical schemes to obtain a self-consistent autocorrelation function of the network noise have been developed (Lerchner et al., 2006; Dummer et al., 2014; Pena et al., 2018), while a general analytical solution to this problem is still an unsolved problem (although van Meegen and Lindner, 2018, developed a theory for the self-consistent autocorrelation of a network of rotators that, via a suitable mapping, can approximate a network of mean-driven integrate-and-fire neurons).

The Amit-Brunel network is the foundation of the model considered in the next chapter, in which a first theoretical description of the single-cell stimulation experiments by Houweling and Brecht (2008) (described in section 1.2) is attempted.

Chapter 2.

Detecting the Stimulation of a Single Cell in a Random Network

The core question raised by the experiment by Houweling and Brecht (2008) can be formulated in just a few words: How can the stimulation of a single cell in the sensory cortex influence the behavior of an animal? However, the chain of events triggered by the stimulation leading to a behavioral response (formation of a perception, decision making, and motor output) possibly involves a very large number of neurons across many regions of the central nervous system so that a "complete" theoretical description of the experiment may have to include a large part - if not all - of the brain. Computational or analytical tractability aside, it is unclear what kind of insight a complicated large-scale model would bring.

As a first step to confine the problem, one could imagine placing the surroundings of the stimulated cell under a spotlight. It is plausible that the initial effects of the stimulation are limited to a localized region. Under what circumstances can these effects make a difference to the rest of the brain? Theoretical (Monteforte and Wolf, 2010) and experimental studies (London et al., 2010) have argued that cortical networks are chaotic. In a chaotic system subject to both intrinsic and external noise sources, it seems likely that reliable encoding must be based on averages over large numbers of neurons. Consistent with this view, it can be postulated that, if the stimulation is to be perceived by the animal, it must elicit a (statistically) significant change in a large population. In this way, the opening question has been modified to: How can single-cell stimulation cause a statistically significant change in the activity of the surrounding network?

Although more limited in scope, this question is still too vague in two regards. First, it is not clear what statistical property of the network's activity should be the relevant one. A possible answer is given later in this chapter. Secondly, what is a reasonable model for the local network? Indubitably, there is no unique answer to this question. However, a first constraint on the model can be posed by taking basic experimental facts into account.

Cortical neurons generally emit only a few spikes per second, at irregular intervals (Shadlen and Newsome, 1998), and in the barrel cortex firing rates are particularly low (Brecht and

Sakmann, 2002; de Kock et al., 2007). Furthermore, neurons in the barrel cortex fire rather asynchronously even during anesthesia (Middleton et al., 2012). Because the detection task demands a high level of attention, which is generally associated to desynchronized firing (Renart et al., 2010; Harris and Thiele, 2011), *a fortiori*, it seems justified to assume the spontaneous activity of the network to be asynchronous and irregular. The "Amit-Brunel network" (see section 1.5), has been the object of analytical studies showing that it can reproduce, depending on the choice of parameters, various firing patterns, including a stable asynchronous irregular state (Amit and Brunel, 1997a; Brunel, 2000). For this reason and because of its comparative simplicity, the Amit-Brunel network seems like a natural starting point for a first approach to the problem.

With the choice of the network model, the central aim of this chapter finally defines itself as the test of something akin to a "null hypothesis": is there a way to detect the single-cell stimulation in a random network of excitatory and inhibitory integrate-and-fire neurons? A second related question is what the optimal conditions (in terms of model parameters) are that permit the detection of the stimulation.

The chapter begins with a description of the network model and the characterization of the spontaneous network activity (section 2.1). The following section 2.2 deals with the effects of the single-cell stimulation on the firing rate of the network. Section 2.3 introduces a detector for the single-cell stimulation and develops a theory to estimate the detection rates analytically. The detector receives input from a readout population that can be biased towards neurons receiving direct input from the stimulated cell. The main result of the chapter is presented in section 2.4: for a sufficiently large bias (representing the effect of the training phase the animals undergo), the single-cell stimulation is detectable with detection rates similar to the experimental ones and, as in the experiments, inhibitory cells are more detectable. If the strength of the recurrent coupling is increased beyond a critical value, however, the detectability deteriorates rapidly. Section 2.5 is concerned with the robustness of the main results with respect to the choice of parameters, with particular emphasis on the size of the network. The last result is presented in section 2.6, which considers a single "barrel". In this case, the analytical calculation of the detection rates yields fairly accurate results without needing any measurement of the network's spontaneous activity. The final section 2.7 offers a summary of the results and discusses limitations.

2.1. Model

The network model consists of $N_E = 80 \cdot 10^3$ excitatory and $N_I = \gamma N_E = 20 \cdot 10^3$ inhibitory leaky integrate-and-fire (LIF) neurons (see section 1.4 on p. 23). The ratio of inhibitory to excitatory neurons has the standard value $\gamma = 1/4$. The total size of the network $N = N_E + N_I = 10^5$ corresponds to about one fourth of the estimated size of the barrel cortex or to an area spanning about five "barrels" (Meyer et al., 2010).

All neurons in the network have identical properties. Excitatory and inhibitory neurons differ only in the effect their spikes have on their targets. Dale's principle is obeyed: neurons are either excitatory or inhibitory. In other words, all entries of a single column in the weight matrix are either positive or negative. Connections are random with fixed in-degree: each neuron receives input from $C_E = 4000$ randomly selected excitatory neurons and $C_I = \gamma C_E = 1000$ randomly selected inhibitory neurons. The total number of inputs per neuron $C = C_E + C_I = 5000$ is in a plausible range for the rat (somatosensory) cortex (Schnepel et al., 2014). Self-connections, also known as autapses, are excluded, which implies that the connection probability between two randomly selected neurons is $p_{c,e} = C_E/(N_E - 1)$ if the presynaptic neuron is excitatory and $p_{c,i} = C_I/(N_I - 1)$ if the presynaptic neuron is inhibitory. However, because $N_E, N_I \gg 1$ the connectivity is almost homogeneous ($p_{c,e} \approx p_{c,i} \approx p_c = C/N$) and sparse ($p_c = 0.05$). Sparse connectivity is often assumed in models of cortical networks and is consistent with the average connection probability between excitatory neurons in barrel cortex (Lefort and Petersen, 2017).

The membrane voltage of the kth neuron evolves according to

$$\tau_m \dot{v}_k = -v_k + R_m[I_{\text{ext}}(t) + I_{\text{syn,k}}(t)]. \tag{2.1}$$

where $\tau_m = 20\,\text{ms}$ is the membrane time constant, $I_{\text{syn,k}}(t)$ the input from the recurrent network, and $I_{\text{ext}}(t)$ models the input from outside the network. Whenever the voltage (measured with respect to the resting potential) $v_k(t)$ reaches $v_T = 20\,\text{mV}$, the neuron fires a spike and $v_k(t)$ is reset to $v_R = 10\,\text{mV}$ after a refractory period $\tau_{\text{ref}} = 2\,\text{ms}$. Delta functions centered on the time of each threshold crossing, $t_{k,l}$, define the output spike train $x_k(t) = \sum_l \delta(t - t_{k,l})$ of the kth neuron. The numerical values for $\tau_m, \tau_{\text{ref}}, v_R, v_T$ are the same as in Brunel (2000) and are on the same order of magnitude of measurements for cortical cells (Beierlein et al., 2000; Harrison et al., 2015).

Neurons are coupled by current-based delayed instantaneous synapses (see section 1.4.1 on p. 23). Let $\{\epsilon_{k,i}\}_{i=1,2\ldots C_E}$ run over the indexes of all excitatory neurons providing input to the kth neuron, i.e. over all C_E non-zero entries of the kth row of the adjacency matrix. Analogously, let $\{\iota_{k,j}\}_{j=1,2\ldots C_I}$ run over all indexes of the inhibitory neurons projecting to neuron k. Furthermore, let J_{nm} and D_{nm} indicate the coupling strength and the transmission delay for the connection from the mth to the nth neuron, respectively. The synaptic input current to the k neuron is

then:

$$I_{\mathrm{syn,k}}(t) = \frac{\tau_m}{R_m}\left[\sum_i^{C_E} J_{k\epsilon_{k,i}} x_{\epsilon_{k,i}}(t - D_{k\epsilon_{k,i}}) - g\sum_j^{C_I} J_{k\iota_{k,j}} x_{\iota_{k,j}}(t - D_{k\iota_{k,j}})\right]. \tag{2.2}$$

The parameter g sets the relative strength of inhibition compared to excitation. The coupling weights J_{nm} are drawn independently from an exponential distribution, which is an approximation to the long-tailed histograms of synaptic efficacies measured in the cortex (Song et al., 2005; Lefort et al., 2009). Random weights form the main difference to the standard Amit-Brunel model and are the principal source of heterogeneity in the network. The mean value of the coupling is indicated with J and has a standard value of $J = 0.1\,\mathrm{mV}$. The strength of the recurrent inhibition g is crucial for the network dynamics and is discussed in detail below. Synaptic transmission delays are uniformly distributed between $D_{\min} = 0.5\,\mathrm{mV}$ and $D_{\max} = 2.0\,\mathrm{mV}$.

As far as the external input is concerned, two scenarios are considered. In the first one, the network is *autonomous*, i.e. it receives no external time-dependent input and $I_{\mathrm{ext},k}(t)$ is simply a constant value:

$$I_{\mathrm{ext},k}(t) = I_0. \tag{2.3}$$

In the second one, the network receives external input and $I_{\mathrm{ext},k}(t)$ the sum of two terms: one constant input I_0 and one time-dependent part, i.e. Poissonian shot-noise mimicking input from other brain areas:

$$I_{\mathrm{ext},k}(t) = I_0 + \frac{\tau_m}{R_m}\left[\sum_{j=1}^{C_{\mathrm{ext}}}\sum_l J_{k,j,l}\delta(t - t_{k,j,l})\right], \tag{2.4}$$

where $t_{k,j,l}$ are independent spiking times with mean rate r_{ext}, C_{ext} is the number of external inputs per neuron and $J_{k,j,l}$ are i.i.d. samples from an exponential distribution. The standard value for the mean amplitude of the shot-noise input is chosen for simplicity to be the same as the mean recurrent coupling $J_{\mathrm{ext}} = 0.1\,\mathrm{mV}$. Anatomical studies indicate that the main origin of feed-forward input drive to barrel cortex is the thalamus (Beierlein et al., 2003; Poulet et al., 2012; Feldmeyer et al., 2013). It has been estimated that only between 5% and 20% of the excitatory input connections to the barrel cortex originate from the thalamus (Schoonover et al., 2014). However, the firing rate of thalamic cells is higher and varies from 5 Hz to 20 Hz depending on the brain state (Voigt et al., 2008; Poulet et al., 2012). Within these ranges, choosing $r_{\mathrm{ext}} = 12\,\mathrm{Hz}$ and $C_{\mathrm{ext}} = 700$ makes the total external input rate similar to the total recurrent excitatory input. On the one hand, this choice of parameters is consistent with several experimental studies highlighting the large impact that thalamic inputs can have on cortical activity (Sun et al., 2006; Poulet et al., 2012); on the other hand, it ensures that the external input is strong but not overwhelming compared to the recurrent network input.

The analysis by Brunel (2000) shows that there are two crucial conditions for a stable asyn-

chronous irregular firing regime.[1] First, the net *recurrent* input from the network must be inhibitory. Second, the mean *external* input must be above the firing threshold. The mean recurrent input from the network must obey

$$\langle R_m I_{\text{syn}}(t) \rangle = \tau_m J C_E (1 - g\gamma) r_{\text{sp}} < 0, \tag{2.5}$$

where the average is here over time and neurons, and r_{sp} is the mean spontaneous firing rate of the network. Fulfilling the first condition requires $g > 1/\gamma$. To meet the second requirement, the constant term I_0 must be chosen such that

$$\langle R_m I_{\text{ext}}(t) \rangle = R_m I_0 + \tau_m J_{\text{ext}} C_{\text{ext}} r_{\text{ext}} > v_T. \tag{2.6}$$

Satisfying the two requirements eqs. (2.5) and (2.6) ensures that the spontaneous activity is asynchronous and irregular. At the same time, consistency with the biological findings demands a low mean spontaneous firing rate. One way to reduce the mean firing rate is to lower the mean external input. However, if $\langle R_m I_{\text{ext}} \rangle$ is lowered, fixed point corresponding to the asynchronous state eventually loses stability via a Hopf bifurcation (Brunel, 2000). A sharp transition only occurs for an infinitely large network, whereas in a finite-sized network the precursors of such bifurcation appear earlier in form of a global oscillation in the network activity. The more $\langle R_m I_{\text{ext}} \rangle$ is reduced, the stronger the global oscillation in the spontaneous activity becomes. The other way to decrease the firing rate, i.e. increasing g, can be exploited up to the point where the synaptic amplitudes become unrealistically strong compared to the excitatory ones. As a compromise, the relative strength of inhibition is set to $g = 7$ and the mean input current I_0 is chosen such that the mean total external input is $\langle R_m I_{\text{ext}}(t) \rangle = 22\,\text{mV}$. In the case of autonomous network, $r_{\text{ext}} = C_{\text{ext}} = 0$, so that $R_m I_0 = 22\,\text{mV}$, while in the presence of external shot-noise the condition is fulfilled by setting $R_m I_0 = 5.2\,\text{mV}$.

With this choice, the average spontaneous firing rate is $r_{\text{sp}} \approx 2\,\text{Hz}$ and the network activity looks rather asynchronous, as seen in the raster plot in fig. 2.1A (the case of autonomous network is shown, the case with external input looks similar). As mentioned above, a perfectly asynchronous state cannot exist in a finite network, because noise fluctuations constantly perturb the system out of the fixed point corresponding to the asynchronous irregular state, thus sustaining the noisy oscillation that can be noticed in the time-dependent firing rate of the entire network.

[1] Although Brunel (2000) did not consider randomly distributed weights, it can be assumed that this one difference will be, in first approximation, equivalent to an increase in the noise and will not change the picture qualitatively, so that his analysis can be used as guide to tune parameters to achieve the desired spontaneous firing regime.

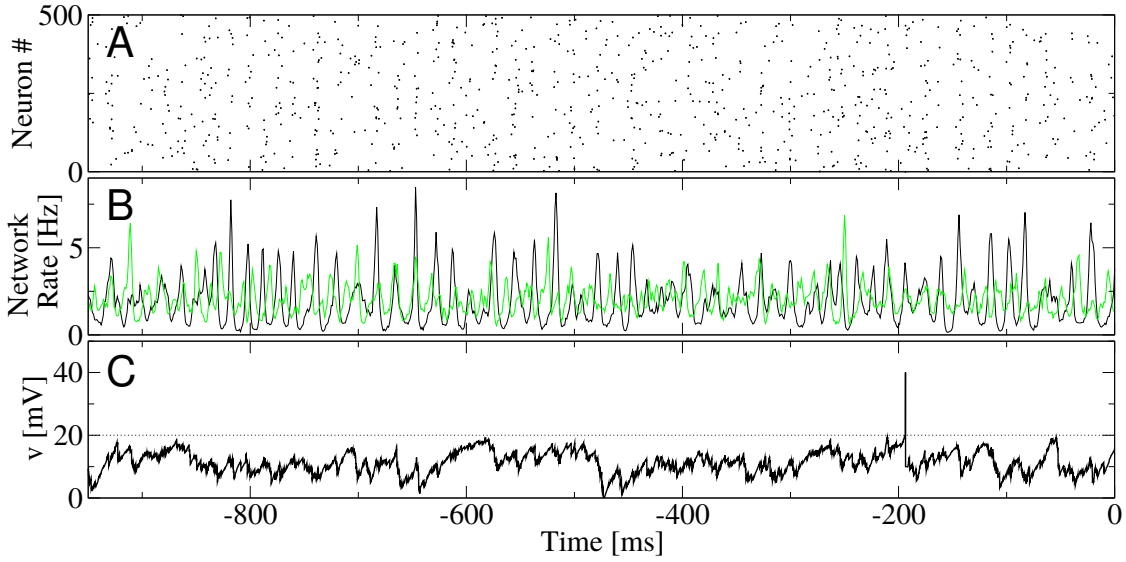

Figure 2.1. – **Visualization of the spontaneous network activity.** Parameters are as in table 2.1 except for the green line in panel **B**. **A**: Raster plot for 500 neurons. **B**: Time-dependent firing rate of the entire network. The slow oscillation with frequency $\approx 60\,\mathrm{Hz}$ is evident. For the case with external shot noise (green line, parameters as in table 2.2) the oscillation is weaker but still present. **C**: Voltage trace of one neuron. The spike has been painted for illustration purposes and does not result from the model dynamics.

The network firing rate is defined as

$$r_{\mathrm{net}}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t), \tag{2.7}$$

and is plotted in fig. 2.1B (here spike trains are convolved with a box of unit area and width $\Delta t = 1\,\mathrm{ms}$ for visualization). The amplitude of the oscillation is somewhat smaller in the presence of external input noise (fig. 2.1B, green line) compared to the autonomous case (fig. 2.1B, black line). Although the network activity is, strictly speaking, not perfectly asynchronous, cross-correlations are weak, so that it makes sense to describe the network state as asynchronous with a finite-size oscillatory perturbation.

Figure 2.1C shows the voltage trace of one neuron, which fluctuates around a value close to the total mean input

$$\langle R_m I(t)\rangle = R_m I_0 + \tau_m J_{\mathrm{ext}} C_{\mathrm{ext}} r_{\mathrm{ext}} + \tau_m J C_E (1 - g\gamma) r_{\mathrm{sp}} \approx 10\,\mathrm{mV}. \tag{2.8}$$

Although the total mean external input is above threshold, eq. (2.8) and fig. 2.1C show that the voltage fluctuates around a value that lies well below the firing threshold. The reason is that

input from the recurrent network is, on average, strongly negative (recall that $\gamma g < 1$). Therefore, if the recurrent network input is shut off, neurons would be driven by the external input and fire towards the threshold (similarly to fig. 1.8A). However, when the negative recurrent feedback from the network is present, the mean of the *total* input is below the firing threshold (as in fig. 1.8B) and neurons can only reach the threshold owing to strong input fluctuations, which occur rarely and at irregular times.

Before further characterizing the spontaneous state of the network, the description of the model needs to be completed with a few technical specifications. Network simulations were implemented in C++ and neurons were integrated with an Euler-method and time step $\Delta t_{\text{sim}} = 0.1\,\text{ms}$. Furthermore, to achieve a sensible improvement in the usage of computational resources, weights were discretized.[2] Initial conditions for voltage, refractoriness and input current were randomly drawn in each trial, whereas the network connectivity (weights and delays) was drawn once and unchanged across trials, therefore playing the role of frozen disorder. To forget initial conditions and reach the stationary state, simulations were run for $400 < T_{\text{ic}} < 1200$ ms before starting data acquisition. The value of T_{ic} was based on the network's autocorrelation time, which is in turn related to the average recurrent coupling (see section 2.8 to for more details).

Numerical values for all parameters used in this chapter are recapitulated in three tables in section 2.8 (p. 96 onwards). Each table refers to a parameter set. Table 2.1 summarizes the parameters for the autonomous network introduced in the present section. Table 2.2 refers to the case with external input. These two parameter sets are referred to as "standard" parameters (with and without external drive) and are used across most of the chapter.

Asynchronous irregular spontaneous network activity

Typically, LIF networks are studied theoretically by approximating the input to each neuron with Gaussian white noise, as briefly outlined in section 1.5.2. This approach neglects both temporal correlations and the shot-noise character of the actual network input. While an analytical self-consistent autocorrelation for LIF networks is still an open question (for numerical schemes see Lerchner et al., 2006; Dummer et al., 2014; Pena et al., 2018), the diffusion approximation can be improved with respect of the shot-noise nature of the input. Richardson and Swarbrick (2010) calculated analytic expressions for a LIF neuron driven by excitatory and inhibitory shot noise with exponentially distributed amplitudes. In particular, if a_e (a_i) and R_e (R_i) are the mean value and the firing rate of the excitatory (inhibitory) shot noise, respectively, the output firing

[2] The most resource-consuming part of a network simulation is the efficient storage of the connection matrix and of the $C \cdot N = 5 \cdot 10^8$ synaptic weights and delays. By combining a sparse matrix representation and using integers instead of floating point numbers, the RAM usage of simulations was reduced to 12 bytes per connection, which is a factor three better than the C++ based NEST package and by a factor ten compared to the Python-based BRIAN. Introducing a cap at $50 \cdot gJ$ for weights and using integers (2^{32} possible values) permits a rather fine weight discretization. The probability of drawing a single weight (in the entire network) exceeding the cap is negligibly small ($\sim 10^{-13}$).

rate of the LIF neuron is given by:

$$\phi_{sn}(a_e, a_i, R_i, I_0) = \left(\tau_{\text{ref}} + \tau_m \int_0^{1/a_e} \frac{\mathrm{d}s}{s} Z_0^{-1}(s) \left[\frac{e^{s\hat{v}_T}}{1 - a_e s} - e^{s\hat{v}_R} \right] \right)^{-1} \tag{2.9}$$

where $\hat{v}_R = v_R - R_m I_0$, $\hat{v}_T = v_T - R_m I_0$, and $Z_0^{-1}(s) = (1 - a_e s)^{\tau_m R_e}(1 + a_i s)^{\tau_m R_i}$. Equation (2.9) is valid for uncorrelated input with random exponential weighs. In the network, each neuron receives temporally correlated inputs with amplitudes chosen from a *fixed sample* from an exponential distribution. However, if the network firing rate is low, the power spectrum of each input spike train will be approximately flat, i.e. similar to a Poisson input. Furthermore, if the number of inputs per neuron is large, the exponential distribution will be sampled well enough that the superposition of all excitatory inputs will look similar to a Poisson process with firing rate $R_e = C_E \cdot r_{\text{sp}}$ and mean amplitude $a_e = J$. Analogously, the total inhibitory input might be treated as a Poisson process of firing rate $R_i = C_I \cdot r_{\text{sp}}$ and mean amplitude $a_i = g \cdot J$. Performing these substitutions in eq. (2.9) and imposing that the output firing rate of a representative neuron has to be equal to the input firing rate yields an equation for r_{sp}:

$$r_{\text{sp}} = \phi_{sn}(J, g \cdot J, C_E \cdot r_{\text{sp}}, C_I \cdot r_{\text{sp}}, I_{\text{ext}}). \tag{2.10}$$

Equation (2.10) can be solved numerically for r_{sp} to predict the spontaneous firing rate of the network.

Figure 2.2A shows r_{sp} as a function of the mean coupling strength J for the autonomous network. The firing rate displays a declining trend because the strength of the recurrent inhibition is proportional to J. The prediction of eq. (2.10), plotted with a continuous line, is rather accurate up to a mean coupling of about 0.2 mV, while the diffusion approximation overestimates the firing rate except when J is very small. When the coupling parameter becomes larger than a critical value, the fluctuations in the input current become stronger and slower, as observed by Ostojic (2014) and explained by Wieland et al. (2015) as an amplification of the slow components of the input noise by the recurrent network. In this firing regime, the long-range temporal correlations violate one main assumption underlying eq. (2.10), namely that the input spike trains are Poissonian, which leads to increasing discrepancies between theory and simulations for larger coupling. The Poissonian approximation for input spike trains breaks down also for very weak couplings. If the mean recurrent input is weak enough, the *total* mean input eq. (2.8) becomes larger than the firing threshold (this is the case for instance for the data point $J = 0.001$ mV in fig. 2.3A). In this situation, the single neuron's firing regime is mean-driven and with a weak noise (because of the weak coupling). Consequently, the output spike train of each neuron is rather regular and can be hardly approximated by a Poisson process. Because the correlation of single spike trains is retained by the summed input (Lindner, 2006), discrepancies with a theory

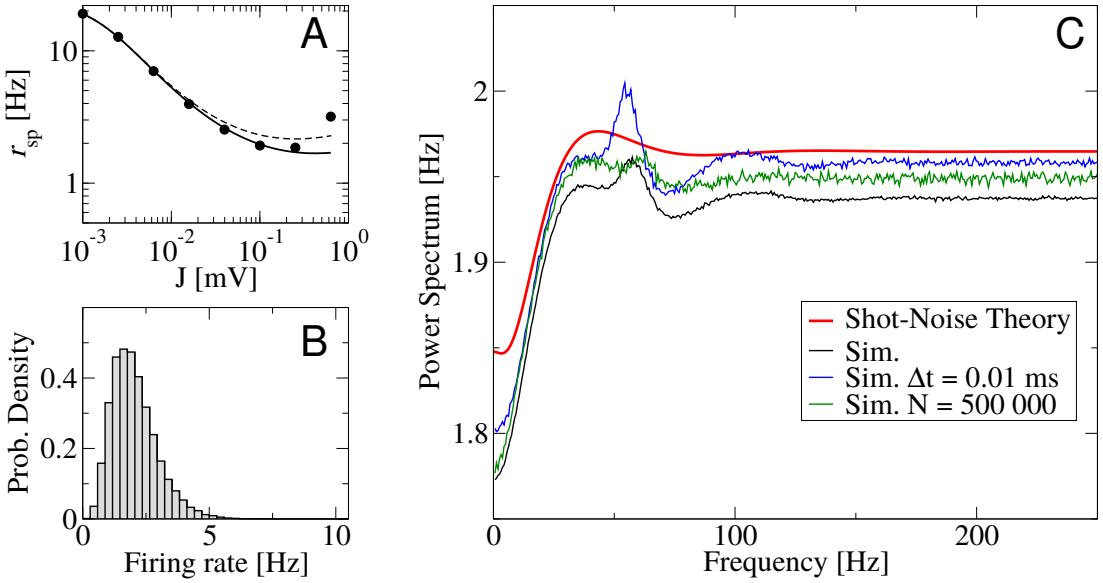

Figure 2.2. – **Statistical properties of the spontaneous network activity.** **A**: Network spontaneous firing rate as a function of the average recurrent coupling J. Continuous line: self-consistent shot-noise theory eq. (2.10), dashed line: self-consistent diffusion approximation, computed as explained in section 1.5.2. **B**: Distribution of firing rates. **C**: Power spectrum averaged over neurons in the network and over runs (with fixed topology). Black line: simulations, standard time step $\Delta t = 0.1\,\text{ms}$; Blue line: time step $\Delta t = 0.01\,\text{ms}$; Green line: standard time step but larger network $N = 5 \cdot 10^5$; Red line: shot-noise theory eq. (2.11). Other parameters as in table 2.1.

assuming Poissonian input could be expected. On the contrary, the spontaneous network firing rate is well predicted by the theory. A likely explanation is that the role of temporal correlations of such a weak noise is marginal compared to that of the mean input, which is still correctly captured by the theory.

The theoretical argument underlying eq. (2.10) assumes a homogeneous network. However, the network considered here is homogeneous only in the statistical sense. In point of fact, each neuron receives a finite number of inputs with different synaptic amplitudes, so that the time-averaged mean input to each neuron is not exactly the same for each neuron, but an (approximately Gaussian) random variable. Consequently, the mean firing rate of each neuron also follows a relatively wide distribution (fig. 2.2B). Note that the firing rate distribution is not Gaussian but skewed to the right because of the nonlinear relation between input and output firing rate (Roxin et al., 2011).

After establishing that the shot-noise theory can be used to approximate the firing rate of neurons in the network, it is natural to ascertain to what extent the theory can also approximate the second-order statistics of the neurons' spike trains, i.e. their power spectrum. To this end, the shot-noise theory by Richardson and Swarbrick (2010) must be adapted to the case of non-

vanishing refractory period, which leads to:[3]

$$S_{xx}(f) = r_{\mathrm{sp}} \left[1 + 2\Re \left\{ \frac{\int_0^{J^{-1}} \mathrm{d}s\, s^{i2\pi f\tau_m} \frac{\mathrm{d}}{\mathrm{d}s} \left[Z_0^{-1}(s) e^{s\hat{v}_R - i2\pi f\tau_{\mathrm{ref}}} \right]}{\int_0^{J^{-1}} \mathrm{d}s\, s^{i2\pi f\tau_m} \frac{\mathrm{d}}{\mathrm{d}s} \left[Z_0^{-1}(s) \left(\frac{e^{s\hat{v}_T}}{1-Js} - e^{s\hat{v}_R - i2\pi f\tau_{\mathrm{ref}}} \right) \right]} \right\} \right]. \tag{2.11}$$

The comparison of eq. (2.11) to the average spectrum measured in the recurrent network is shown in fig. 2.2C (the black line represents simulation results for the standard parameters, the theory is plotted in red), again focusing on the case of the autonomous network. The approximation for the standard parameters is not bad, considering that the input in the network is neither temporally nor spatially uncorrelated, as assumed in the theory. A further source of discrepancies are the time and weight discretization used for network simulations. A shorter simulation time step (fig. 2.2C, blue line) improves the precision of the firing rate, but the shape remains similar, and the peak around 60 Hz with its harmonics become even more evident. As pointed out above, this oscillation is related to the finite size of the network (Brunel, 2000); for a larger, sparser network (fig. 2.2C, green line) the peak is indeed much smaller. Neither the finite time step nor the size of the network have much influence on the low-frequency limit of the spectrum, which is related to the temporal correlations between inter-spike intervals (Cox and Lewis, 1966).

As a final remark, the network average firing rate r_{sp} of a particular realization of the network is also a random variable. As mentioned above and for reasons discussed in the following section, all quantities in this chapter were averaged over trials and (when applicable) over neurons of one network, but *not* over different networks. Hence, also the spectra displayed in fig. 2.2C were obtained by averaging over trials and neurons of one particular network. The realization shown in fig. 2.2 has a r_{sp} which is below its average over network realizations. Therefore, the agreement between shot-noise theory and the power spectrum averaged over networks is slightly better (not shown).

[3] The term in curly brackets in eq. (2.11) is the Fourier transform of the spike-triggered rate $\rho(t)$ and is obtained by solving the Fourier transform (both in time and voltage) of the master equation

$$\partial_t P + \partial_v J = \rho(t - \tau_{\mathrm{ref}})\delta(v - v_R) - \rho(t)\delta(v - v_T) + \delta(t - \tau_{\mathrm{ref}})\delta(v - v_R),$$

where P and J are the probability density and flux, respectively. The calculation is not reported here because it is essentially the same as in the Supplementary Material of the paper by Richardson and Swarbrick (2010) with an additional factor due to the non-zero refractory period.

2.2. Single-cell stimulation and firing-rate response

To model the single-cell stimulation experiment, the network is simulated for a total time window of $T = 3\,\mathrm{s}$ centered on $t = 0$. A randomly selected neuron is chosen and labeled as $\mathcal{B}_0$. For $0 < t < T_s = 400\,\mathrm{ms}$ the constant input of $\mathcal{B}_0$ is then increased by $R_m \Delta I_{\mathrm{ext}} = 23\,\mathrm{mV}$ to bring its firing rate from the spontaneous value r_{sp} to a new value r_0. In each trial, initial conditions are randomly varied; however, the network realization and the choice of $\mathcal{B}_0$ are not. Changing either $\mathcal{B}_0$ or the network connectivity would be tantamount to changing cell or animal. In the experiments, the same cell is used for many trials, but the final effect size results from an average over many cells. Here, for simplicity the network realization is frozen. One way to partially represent the effect of changing cells without changing network is discussed later on.

Although the network topology is homogeneous, stimulating $\mathcal{B}_0$ "breaks the symmetry" so that the effects on other neurons are not homogeneous. In particular, two subsets of the network (excluding $\mathcal{B}_0$ itself) must be distinguished: the set of neurons receiving direct input from $\mathcal{B}_0$, labeled as $\mathcal{B}_1$, and the set of all other neurons, labeled as $\mathcal{B}_2$. From this definition, it follows that neurons belonging to $\mathcal{B}_1$ are one link away from $\mathcal{B}_0$, while neurons belonging to $\mathcal{B}_2$ are at least two links away. For the parameters considered here, it turns out that all neurons in $\mathcal{B}_2$ are exactly two links away from $\mathcal{B}_0$, because the probability for a neuron to be three synapses away from $\mathcal{B}_0$ (or from any given neuron) is extremely small.[4] These subsets are also depicted in fig. 2.3A (note that there is no spatial structure in the network and neurons are grouped only for the ease of illustration).

The time-dependent firing rate of each subpopulation $\mathcal{B}_k$ ($k = 0, 1, 2$) is defined as (angular brackets indicate trial-average and N_k denotes the size of $\mathcal{B}_k$)

$$r_k(t) = \left\langle \frac{1}{N_k} \sum_{x \in \mathcal{B}_k} x(t) \right\rangle . \tag{2.12}$$

The firing-rate response discussed in the following of this section is defined as the deviation of $r_k(t)$ from the spontaneous value:

$$\Delta r_k(t) = r_k(t) - r_{\mathrm{sp}}. \tag{2.13}$$

During the stimulation, the system quickly reaches a new fixed point, as shown later on in this section. To avoid the proliferation of symbols, the new steady-state values of $r_k(t)$ will be indicated as r_k, i.e. the same symbol without time argument. Analogously, Δr_k indicates the

[4] The probability for a neuron to be (at least) three synapses away from $\mathcal{B}_0$, p_3, is the probability that no direct connection exists from *all* neurons in $\mathcal{B}_1$ to the target neuron. Because there are on average $N_1 = 5000$ neurons in $\mathcal{B}_1$ (the probability distribution of N_1 is binomial with a narrow relative standard deviation), it results that $p_3 \approx (1 - p_c)^{N_1} \sim 10^{-112}$. The probability of finding at least one neuron in $\mathcal{B}_2$ that is more than three synapses away from $\mathcal{B}_0$ is then $1 - (1 - p_3)^{N_2} \approx N_2 p_3 \sim 10^{-107}$, where N_2 is the size of $\mathcal{B}_2$.

steady-state value of eq. (2.13). A theoretical prediction for r_0, r_1, and r_2 can be obtained, similarly to the spontaneous firing rate, by imposing self-consistency between input and output firing rates of all three subpopulations. Let $\nu_{e,k}$ be the total input excitatory rate of subpopulation $\mathcal{B}_k$ ($k = 0, 1, 2$) and with $\nu_{i,k}$ the total input inhibitory rate of $\mathcal{B}_k$. These input rates are given by the sum of the input rates from all other subpopulations multiplied by the respective average number of input connections to $\mathcal{B}_k$. In the case that $\mathcal{B}_0$ is excitatory, the new steady-state firing rates of the autonomous network are found by solving the following system:

$$\begin{aligned}
r_0 &= \phi_{sn}(\nu_{e,0}, \nu_{i,0}, I_{\text{ext}} + \Delta I_{\text{ext}}) \\
r_1 &= \phi_{sn}(\nu_{e,1}, \nu_{i,1}, I_{\text{ext}}) \\
r_2 &= \phi_{sn}(\nu_{e,2}, \nu_{i,2}, I_{\text{ext}}) \\
\nu_{e,0} &= p_c C_E r_1 + (1 - p_c) C_E r_2 \\
\nu_{e,1} &= r_0 + p_c(C_E - 1) r_1 + (1 - p_c)(C_E - 1) r_2 \\
\nu_{e,2} &= p_c C_E r_1 + (1 - p_c) C_E r_2 \\
\nu_{i,0} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2 \\
\nu_{i,1} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2 \\
\nu_{i,2} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2.
\end{aligned} \tag{2.14}$$

In the last equations, the output firing rate of each subpopulation is approximated by the shot-noise theory eq. (2.9) (the first two arguments of ϕ_{sn} have been omitted for simplicity). Furthermore, it is assumed that $(C_E - 1)/(N_E - 1) \approx C_E/(N_E - 1) \approx C_E/N_E = p_c$, i.e. correction terms of order $1/N$ have been neglected. Within this approximation, $\mathcal{B}_1$ contains on average $p_c N_E = C_E$ excitatory neurons. Therefore, the C_E inputs to $\mathcal{B}_0$ and to $\mathcal{B}_2$ originate from $\mathcal{B}_1$ with probability $C_E/N_E = p_c$, from $\mathcal{B}_2$ with probability $1 - p_c$, and from $\mathcal{B}_0$ with probability zero (there are neither direct connections from $\mathcal{B}_0$ to $\mathcal{B}_2$ by definition, nor from $\mathcal{B}_0$ to itself). These considerations explain why the prefactors multiplying r_0, r_1, and r_2 in the fourth and sixth equation are zero, $p_c C_E$, and $(1 - p_c)C_E$, respectively. Excitatory input rates to neurons in $\mathcal{B}_1$ (fifth line in the above system of equations) must be treated differently if $\mathcal{B}_0$ is excitatory. In this case, by definition of $\mathcal{B}_1$, the probability of receiving input from $\mathcal{B}_0$ to $\mathcal{B}_1$ is unity. Consequently, there are $C_E - 1$ input connections left to assign to $\mathcal{B}_1$ and $\mathcal{B}_2$ with probability, as before, p_c and $1 - p_c$, respectively. Because $\mathcal{B}_1$ contains on average $p_c N_I = C_I$ inhibitory neurons, the probability of receiving an inhibitory input from $\mathcal{B}_1$ is $C_I/N_I = p_c$, i.e. the same as for excitatory inputs. Therefore, the $C_I = \gamma C_E$ inhibitory inputs to any cell in the network stem from $\mathcal{B}_1$ and from $\mathcal{B}_2$ in proportions p_c and $1 - p_c$, respectively (there can be no inhibitory input from $\mathcal{B}_0$ by definition, if $\mathcal{B}_0$ is excitatory).

In the case of inhibitory $\mathcal{B}_0$, the above considerations for the term $\nu_{e,1}$ apply to $\nu_{i,1}$, and vice

versa. Accordingly, only two equations change compared to eq. (2.14):

$$\begin{aligned}
r_0 &= \phi_{sn}(\nu_{e,0}, \nu_{i,0}, I_{\text{ext}} + \Delta I_{\text{ext}}) \\
r_1 &= \phi_{sn}(\nu_{e,1}, \nu_{i,1}, I_{\text{ext}}) \\
r_2 &= \phi_{sn}(\nu_{e,2}, \nu_{i,2}, I_{\text{ext}}) \\
\nu_{e,0} &= p_c C_E r_1 + (1 - p_c) C_E r_2 \\
\nu_{e,1} &= p_c C_E r_1 + (1 - p_c) C_E r_2 \\
\nu_{e,2} &= p_c C_E r_1 + (1 - p_c) C_E r_2 \\
\nu_{i,0} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2 \\
\nu_{i,1} &= r_0 + p_c(\gamma C_E - 1) r_1 + (1 - p_c)(\gamma C_E - 1) r_2 \\
\nu_{i,2} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2,
\end{aligned} \tag{2.15}$$

The two systems eqs. (2.14) and (2.15) pertain to the case of autonomous network. Extending the description to the scenario with external shot-noise is straightforward only if the amplitude of the external noise is equal to the mean excitatory coupling, i.e. if $J = J_{\text{ext}}$. In this case, it suffices to add a term $C_{\text{ext}} r_{\text{ext}}$ to the three equations for $\nu_{e,k}$. In the remainder of this section, the firing-rate response of the network will be investigated only for the autonomous network.

The predictions of eqs. (2.14) and (2.15) can now be tested for different values of the mean recurrent coupling J. To this end, the relative steady-state firing-rated deviations in response to the stimulation $\Delta r_k / r_{\text{sp}} = \Delta r_k - r_{\text{sp}} / r_{\text{sp}}$ as a function of the average recurrent coupling J are plotted in fig. 2.3B-D.

The relative firing-rate deviation for $\mathcal{B}_0$ (fig. 2.3B) displays a maximum around $J = 0.1\,\text{mV}$. For decreasing J, it becomes smaller only because of the increasing spontaneous firing rate. For increasing J, it drops. The reason for this decrease is found in the appearance of slow and strong fluctuations in the network noise when the average coupling enters a critical range. As already mentioned in section 2.1, this phenomenon has been explained by Wieland et al. (2015) as an instability in the linear map describing the input-output relationship for the low-frequency limit of the spike train power spectrum. For large values of J, the strong and slow spontaneous fluctuations of the network noise eventually drown the current step. The theoretical prediction is in good agreement with simulations except for the strongest coupling. As already argued when discussing the discrepancy in the theoretical prediction for the spontaneous firing rate (fig. 2.2A), for values of J above the critical point the slow fluctuations in the input rates are in marked contrast with the assumption of temporally uncorrelated inputs underlying the shot-noise theory.

Intuitively, the firing rate of neurons in $\mathcal{B}_1$ rises when $\mathcal{B}_0$ is excitatory and sinks when $\mathcal{B}_0$ is inhibitory, as seen in fig. 2.3C. Here, the relative deviation $\Delta r_1 / r_{\text{sp}}$ is shown as a function

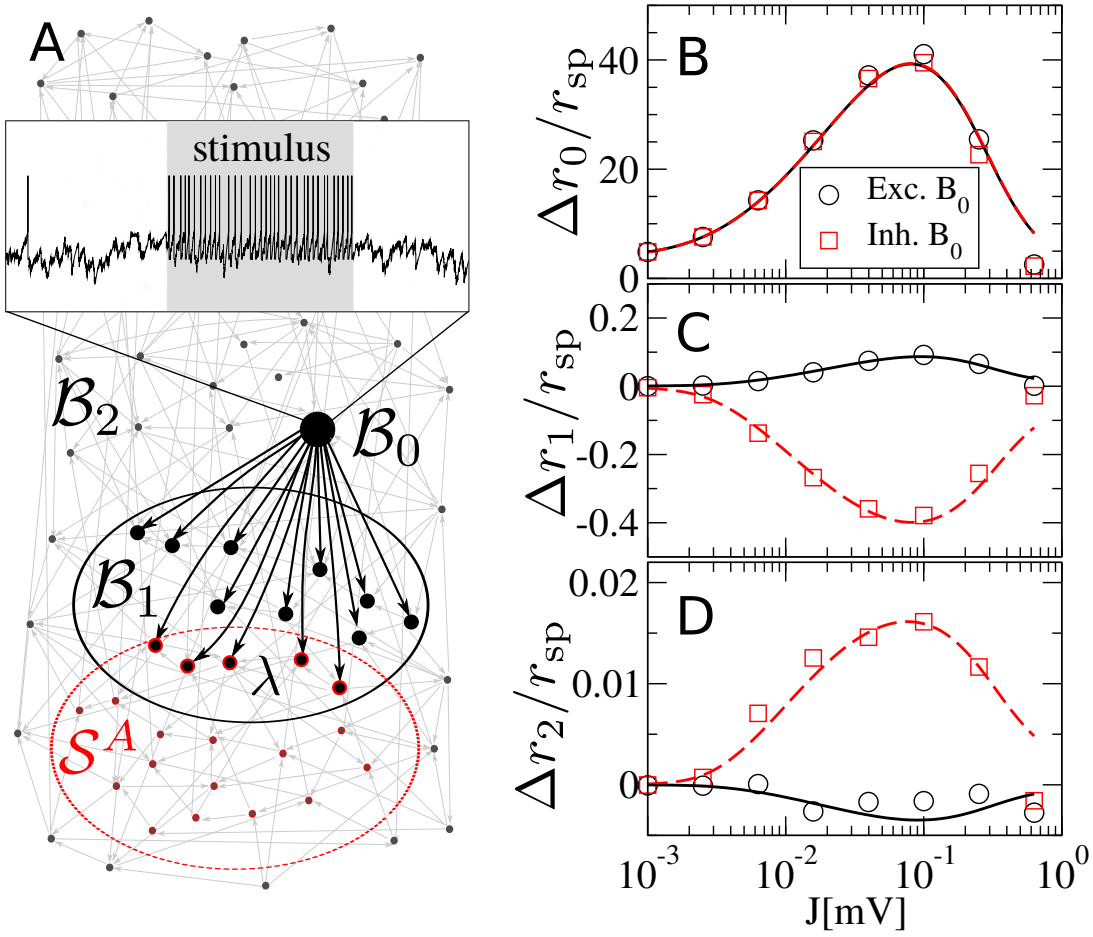

Figure 2.3. – Network model and maximum relative deviation of firing rate for the different subpopulations. **A**: Illustration of the network and notation: $\mathcal{B}_0$ is the stimulated neuron, chosen at random; $\mathcal{B}_1$ is the set of neurons receiving direct connections from $\mathcal{B}_0$; $\mathcal{B}_2$ are all other neurons in the network; the readout population $\mathcal{S}^A$ is introduced at the beginning of section 2.3; note that the network has no structure and neurons are grouped only for illustration convenience. **B**,**C**,**D**: relative firing rate deviation for $\mathcal{B}_0$, $\mathcal{B}_1$, and $\mathcal{B}_2$, respectively. The theoretical prediction is rather accurate for $J \leq 0.3\,\mathrm{mV}$.

of J. The case of excitatory (inhibitory) $\mathcal{B}_0$ is plotted with a continuous (dashed) line for theory and circles (squares) for simulation results. The deviation from the mean is again largest for $J \approx 0.1\,\mathrm{mV}$. The origin of this maximum, as well as the good agreement of theory with simulations except for the largest J, can be explained with the same arguments used for $\mathcal{B}_0$ in the previous paragraph.

Because neurons in $\mathcal{B}_2$ do not receive direct input from $\mathcal{B}_0$, the effect of the stimulus on r_2 can be expected to be weaker. Indeed, $\Delta r_2/r_{\mathrm{sp}}$ is much smaller than $\Delta r_1/r_{\mathrm{sp}}$, as seen in fig. 2.3D (the meaning of symbols and colors is the same as in the previous plot). Less intuitively, the firing-rate response for $\mathcal{B}_2$ is of opposite sign: an excitatory perturbation causes the firing rate of $\mathcal{B}_2$ to decrease, and the other way around for inhibitory $\mathcal{B}_0$. The reason can be understood by realizing that the net input from the network is inhibition-dominated. In other words, if the firing rate of $\mathcal{B}_1$ grows, the net inhibitory input from $\mathcal{B}_1$ to $\mathcal{B}_2$ increases, thus reducing the firing rate of $\mathcal{B}_2$. If the firing rate of $\mathcal{B}_1$ is reduced by an inhibitory $\mathcal{B}_0$, the overall negative recurrent input from $\mathcal{B}_1$ to $\mathcal{B}_2$ decreases, which leads to a higher firing rate within $\mathcal{B}_2$.

As a final remark to the agreement between the theory and simulations for Δr_k, it was pointed out at the end of section 2.1 that the value of r_{sp} depends on many factors neglected in the theory (such as the simulation time step, the sparsity of connections, and even the particular realization of the network topology) that influence the agreement of the measured r_{sp} with the theoretical approximation. However, the same factors have a similar influence also on r_0, r_1, and r_2, so that the discrepancy between theory and simulations mostly cancels out in the difference $r_k - r_{\mathrm{sp}}$ and the agreement between theory and simulation for firing rate deviations is generally much better than for the absolute values of the firing rates.

A thorough theoretical description of the time-dependent firing rates deviations is a more difficult problem. However, if the perturbation is weak compared to the background noise level, a "quasi-stationary" description can provide a reasonable approximation (see for instance Gerstner et al., 2014, chapter 15). In this picture, the time-course of the network firing rate is described by an exponential relaxation from the previous fixed point to the new one, and the time constant can be roughly approximated by the membrane time constant of the neuron. Applying this approach yields:

$$\begin{aligned} \Delta r_k(t) &= r_k(t) - r_{\mathrm{sp}} \approx (r_k - r_{\mathrm{sp}})\Delta a(t) \\ &\approx \Delta r_k[H(t)(1 - e^{-\frac{t}{\tau_m}}) - H(t - T_s)(1 - e^{-\frac{t-T_s}{\tau_m}})], \end{aligned} \tag{2.16}$$

where $H(t)$ is the Heaviside step function. The approximation in eq. (2.16) is compared to simulations in fig. 2.4 in the case that $\mathcal{B}_0$ is inhibitory and the recurrent coupling has the standard value $J = 0.1\,\mathrm{mV}$. The response of $\mathcal{B}_0$ measured from simulations is shown in fig. 2.4A (black line). It is evident that the actual response is much faster than predicted by the theory (red line). This mismatch is not surprising because an instantaneous increase in the input current

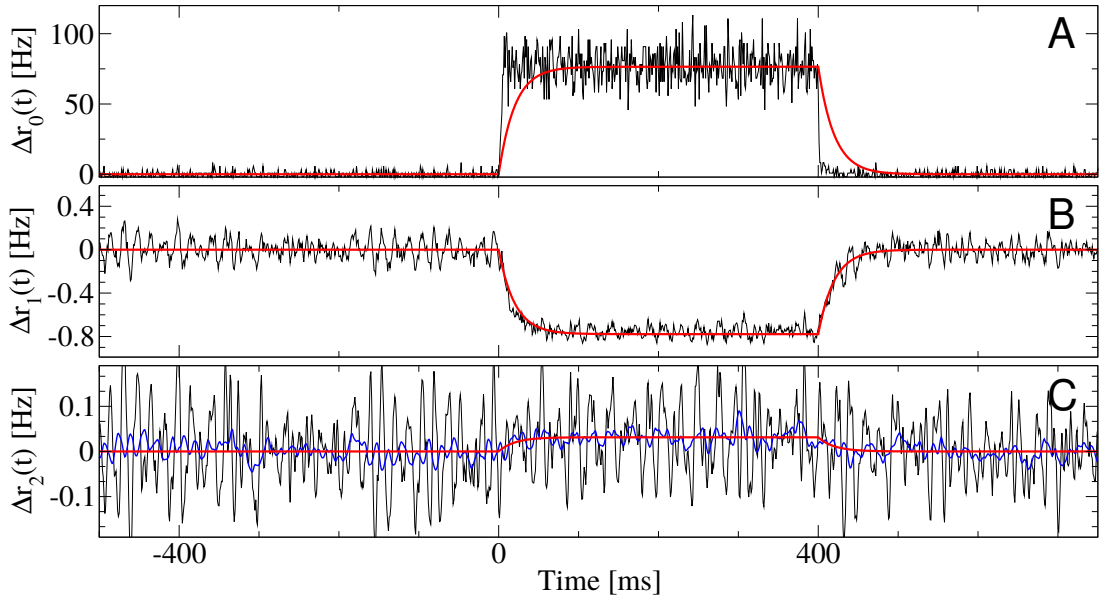

Figure 2.4. – **Time-dependent firing-rate response of the three subpopulations of fig. 2.3.** Here, the case of inhibitory $\mathcal{B}_0$ is shown. The population activity eq. (2.12) was filtered with a box of unit area and width 1 ms (black line) or 20 ms (blue line). The theory (red line) is calculated from eqs. (2.15) and (2.16). **A**: Time-dependent firing-rate deviation from spontaneous value of $\mathcal{B}_0$. **B**: Same for $\mathcal{B}_1$. **C**: Same for $\mathcal{B}_2$. Note that the discrepancies between theory and simulations due to the finite size effects, dependence on network realization etc. mostly cancel out in the difference firing rate deviations $r_k - r_{\rm sp}$. Parameters as in table 2.1.

by more than one hundred percent can hardly be regarded as a weak stimulus (as required by the quasi-stationary approximation). However, the time course of $\Delta r_1(t)$, plotted in fig. 2.4B (black line), and the time course of $\Delta r_2(t)$, shown in fig. 2.4C (black line), are in quite good agreement with the theory (red). To reduce fluctuations and ease the comparison with the theory, in fig. 2.4C the firing rate of $\mathcal{B}_2$ was additionally filtered with a larger time step and plotted in blue.

2.3. Perturbation detection, definitions and theory

The firing-rate response considered in the last section is based on the average over multiple trials. In the experiment, the animals must report trial by trial whether the nano-stimulation was switched on or not. Hence, the model needs to be equipped with a detector that decides on the presence of single-cell stimulation in each trial. Introducing a possible detector is the first goal of this section. Afterwards, a theory is developed to estimate the detection rates analytically and to relate the detectability of the stimulation with the properties of the network.

2.3.1. Readout activity and detector

It is plausible to assume that a neural circuit reading out the activity of the stimulated network cannot access every single neuron in it, but only a subset. In the simplest scenario, a downstream reaction is provoked whenever the activity of this readout subset differs significantly from the spontaneous state. The magnitude of the deviation necessary to cause the reaction depends on the sensitivity of the detector. This idea is made more precise in the following.

The readout set is given the name $\mathcal{S}^A$. If the detector is thought of as a neuron (or a group of neurons), it is natural to set the size of $\mathcal{S}^A$ equal to the number of input connections per neuron C. To be conservative, $\mathcal{B}_0$ is excluded from the readout set. Otherwise, $\mathcal{S}^A$ is formed by choosing randomly λC neurons from $\mathcal{B}_1$ and $(1-\lambda)C$ neurons from $\mathcal{B}_2$ (the readout set is shown in red in fig. 2.3A).[5] By this construction, λ is the prescribed overlap between $\mathcal{S}^A$ and $\mathcal{B}_1$ and quantifies to what extent and in which direction the readout is biased: $\lambda = 1$ indicates the maximum bias *towards* $\mathcal{B}_1$ (perfect overlap between $\mathcal{S}^A$ and $\mathcal{B}_1$), $\lambda = 0$ corresponds to the maximum bias *against* $\mathcal{B}_1$ (neurons in $\mathcal{B}_1$ are excluded from the readout), and the unbiased case is obtained by setting the overlap to its "natural" value $\lambda = \lambda_0$, where $\lambda_0 = C/N = p_c$.

In the experiment, the ability to detect the nanostimulation is gained through a training phase. This learning process implies that something changes in the recurrent connections of the stimulated network or in the outgoing projections to the readout. Because λ can also be interpreted as a bias in the connection probability from $\mathcal{B}_1$ to the detector, prescribing a value of λ different from λ_0 can be regarded as a simple caricature for the learning phase.

One simple possibility to define the readout activity is to consider a smoothed sum of all spike trains in $\mathcal{S}^A$:

$$R^A_\lambda(t) = \frac{1}{C} \sum_{x_i \in \mathcal{S}^A} x_i \star \mathcal{F}_{\tau_f}(t), \tag{2.17}$$

[5] Because N_1 (the size of $\mathcal{B}_1$) is binomially distributed with mean C, there is good chance that $N_1 < C$. In this case, for $\lambda > N_1/C$ there are not enough neurons in $\mathcal{B}_1$ to choose from and the readout set is filled in with neurons from $\mathcal{B}_2$. Because the relative standard deviation of N_1 is small, this situation occurs only for values of λ very close to unity.

where $\star$ indicates convolution and $\mathcal{F}_{\tau_f}$ is a truncated Gaussian filter:

$$\mathcal{F}_{\tau_f}(t) = \frac{1}{\sqrt{(\pi\tau_f^2/2)}} e^{\frac{-(t-3\tau_f/2)^2}{\tau_f^2/2}} H(t)H(3\tau_f - t). \tag{2.18}$$

The filter is causal and its width is set by the parameter $\tau_f = 100\,\text{ms}$, which is twice the standard deviation and $\approx 90\%$ of the full width at half maximum. This time scale is within the experimentally measured range for slow excitatory NMDA synapses, which is 50 ms to 200 ms (Flint et al., 1997). The effect of changing τ_f is discussed in the last part of section 2.5.

As an example, two trials of $R_\lambda^A(t)$ for three values of λ are shown in different colors in fig. 2.5. In each trial, $R_\lambda^A(t)$ fluctuates around its trial-average $\left\langle R_\lambda^A(t) \right\rangle$ (dashed lines of the same color). For $t < 0$, the trial-average does not depend on λ and is equal to the spontaneous value

$$R_{\text{sp}}^A = \frac{1}{T_w} \int_{-T_w}^{0} \mathrm{d}t \langle R_\lambda^A(t) \rangle \approx r_{\text{sp}} \star \mathcal{F}_{\tau_f}(t) = r_{\text{sp}} \int_{-\infty}^{+\infty} \mathrm{d}t\, \mathcal{F}_{\tau_f}(t) \approx r_{\text{sp}}. \tag{2.19}$$

The range of the first integral appearing in eq. (2.19) is in fact arbitrary, as long as the system is in the stationary state (which is the case for $t < 0$ and if enough time has elapsed since the beginning of the simulation). For reasons of symmetry, the range is set at $T_w = 1300\,\text{ms}$, the time window for the perturbation detection. Note that R_{sp}^A is not exactly equal to r_{sp}. One difference is seen in the second approximation in eq. (2.19): the filter is not perfectly normalized to unity. The other difference is due to the heterogeneity of firing rates in the network (see fig. 2.2B): despite the large size of the readout set, the spontaneous firing rate averaged over neurons within $\mathcal{S}^A$ slightly depends on the particular choice of neurons forming $\mathcal{S}^A$. In other words, each readout set has its own R_{sp}^A. After the stimulus onset, the trial average settles transiently on a value approximately equal to $r_\lambda = r_1\lambda + r_2(1-\lambda)$. This temporary plateau of the readout activity can either be above or below the spontaneous value, depending on the type of the stimulated cell and the value of λ.

The task of a detector is to react when the activity of the readout set deviates from the spontaneous state by an amount exceeding its sensitivity. As observed in the last paragraph, the sign of the average deviation depends on λ and on the type of the stimulated cell. Therefore, if the detector is to be impartial about the identity of $\mathcal{B}_0$, it must treat upward and downward deflections of the activity equally. In practice, the detector can evaluate the absolute value of the deviation of the readout activity from the spontaneous state

$$\left|\Delta R_\lambda^A(t)\right| = \left|R_\lambda^A(t) - R_{\text{sp}}^A\right|, \tag{2.20}$$

and respond whenever $\left|\Delta R_\lambda^A(t)\right|$ exceeds a threshold value θ_d. In other words, a reaction is

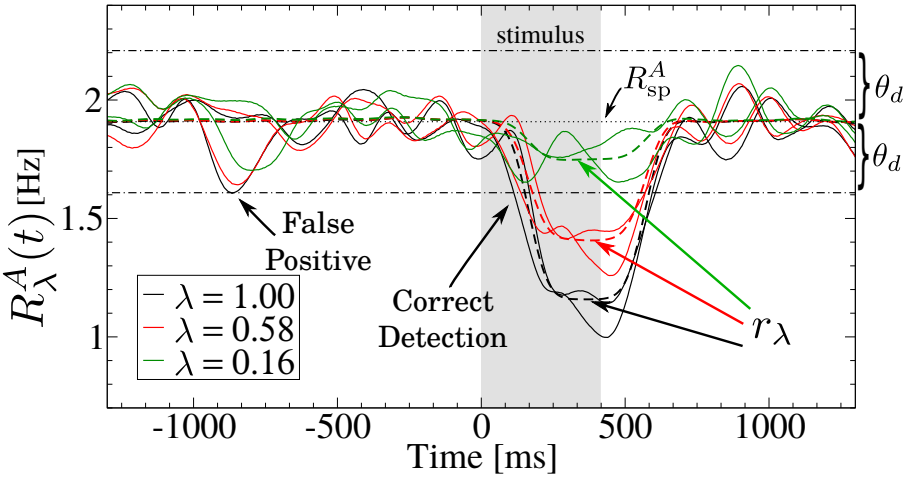

Figure 2.5. – **Readout activity and stimulus detection.** Two example realizations of the readout activity $R^A_\lambda(t)$ are plotted as continuous lines for three values of the bias λ (the color refers to the value of λ as shown in the legend). The trial average corresponding to each λ, $\langle R^A_\lambda(t)\rangle$, plotted with dashed lines, reaches a plateau $\approx r_\lambda$. Crossings of either of the two barriers placed at $R^A_{\rm sp} \pm \theta_d$ define *hits* or *correct detections* if they occur in the detection window placed after the stimulus onset, i.e. for $t \in (0, T_w)$. If the $R^A_\lambda(t)$ crosses one of the two barriers in the detection window before the stimulus onset, i.e. for $t \in (-T_w, 0)$, a *false positive* is registered. The sensitivity of the detector is set by the parameter θ_d, i.e. the distance of the two barriers from the time-average of $R^A_\lambda(t)$ in the absence of stimulus. Parameters are as in table 2.1 with inhibitory $\mathcal{B}_0$.

triggered whenever $R^A_\lambda(t)$ leaves the interval $\left(R^A_{\rm sp} - \theta_d, R^A_{\rm sp} + \theta_d\right)$; if this happens before the stimulus onset (for $-T_w < t < 0$), this defines as *false positive* event. If the detection event occurs after the stimulus onset (but within the detection time window, i.e. $0 < t < T_w$), a *correct detection* event is registered. An example for both events is shown in fig. 2.5. The fraction of trials in which a false positive is detected defines the false positive rate as a function of the detection threshold (the meaning of the outer angular brackets is explained after the next definition):

$$\mathcal{FP}_\lambda(\theta_d) = \left\langle \left\langle \max_{t\in(-T_w,0)} \left\{ H(|\Delta R^A_\lambda(t)| - \theta_d) \right\} \right\rangle \right\rangle_{\mathcal{S}^A}. \tag{2.21}$$

The hit (or correct detection) rate $\mathcal{CD}(\theta_d)$ is defined analogously but for a threshold crossing in $0 < t < T_w$:

$$\mathcal{CD}_\lambda(\theta_d) = \left\langle \left\langle \max_{t\in(0,T_w)} \left\{ H(|\Delta R^A_\lambda(t)| - \theta_d) \right\} \right\rangle \right\rangle_{\mathcal{S}^A}. \tag{2.22}$$

Because both hit and false positive rates can depend on the particular realization of the readout set $\mathcal{S}^A$, they were both averaged over sixteen realizations of $\mathcal{S}^A$, which is indicated by the outer angular brackets in the two last equations.

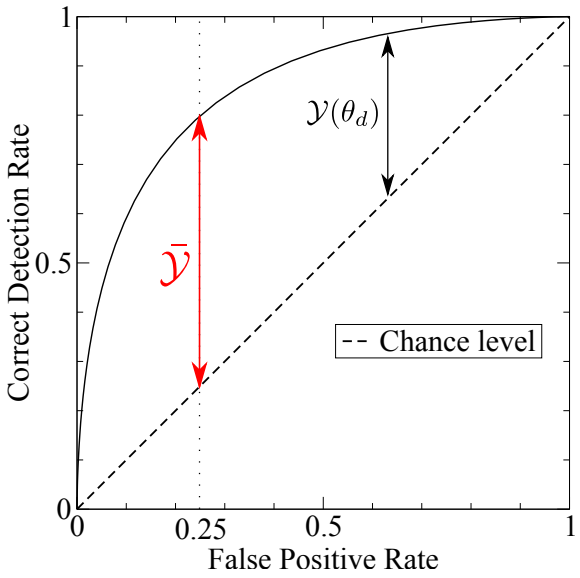

Figure 2.6. – Receiver operating characteristic (ROC) curve and effect size $\bar{\mathcal{Y}}$.

Plotting the correct detection rate as a function of the false positive rate upon variation of θ_d defines the so-called *receiver operating characteristic* (ROC) curve, which is a canonical way of representing the performance of a detector (fig. 2.6). The diagonal represents the chance level, i.e. a useless detector that reacts to signals with the same probability as to the absence of a signal, the ideal detector is the upper edge of the plot, and real detectors are in-between these two extremes.

Houweling and Brecht (2008) expressed their main result as the difference between hit and false positive rates, which they termed *effect size*. In our model, the difference between hit and false positive rate

$$\mathcal{Y}_\lambda(\theta_d) = \mathcal{CD}_\lambda(\theta_d) - \mathcal{FP}_\lambda(\theta_d) \tag{2.23}$$

depends on θ_d and is represented by the vertical distance from the ROC curve to the diagonal. The position of the threshold θ_d determines the detector's sensitivity: a high θ_d corresponds to a low sensitivity, whereas a low θ_d increases the responsiveness of the detector, causing both more false positives and correct detection events. However, in the experiments the responsiveness cannot be directly controlled and is just an attribute of each animal. The false positive rates measured experimentally are scattered over a rather broad range (0 to 0.5) with average just below 0.25. For simplicity, the threshold $\bar{\theta}_d$ that yields a false positive rate of 25% will be chosen in the following

$$\mathcal{FP}_\lambda(\bar{\theta}_d) = 0.25. \tag{2.24}$$

Inserting $\bar{\theta}_d$ into eq. (2.23) yields an effect size that does not depend on θ_d and can be compared

to experiments (fig. 2.6 red arrow):

$$\bar{\mathcal{Y}}(\lambda) = \mathcal{Y}_\lambda(\bar{\theta}_d). \tag{2.25}$$

One potential issue with the definition in eq. (2.25) is that the maximal effect size is not 100%, but only 75%. In practice, however, this intrinsic cap is not relevant because it is well above the largest effect size measured for any cell. In fact, the average effect size measured in the experiments is rather small, so that computing its statistical significance was essential. Here, both to be consistent with the experiments and because simulating large networks is computationally demanding, results are based on a number of trials similar to the experimental ones. Consequently, assessing the statistical significance of the model results is equally important.

For any given value of the detection threshold and for a given realization of the readout set $\mathcal{S}^A$, our virtual detection experiment is a *binary classification problem*. The results for such an experiment are usually represented in a so-called *contingency table*, which registers the occurrences of the four possible outcomes: the number of hits (correct detections), misses, false positives, and correct rejections. The usual statistical test for this case is Fisher's exact test, which assigns a p-value to each contingency table. This p-value represents the probability that the observed contingency table, or a more unlikely one, result out of chance, assuming that the null hypothesis is true. The null hypothesis is that the presence of the stimulus and the reaction are independent of each other. However, averaging over multiple realizations of the readout set adds a complication: each $\mathcal{S}^A$ yields its own contingency table. From each of these tables, one effect size and one p-value could be obtained. Although the effect sizes can surely be averaged, averaging p-values does not make much sense. Importantly, the p-values resulting from each realization of $\mathcal{S}^A$ set are *not* independent, because activities corresponding from multiple readout sets are correlated by global fluctuations in the network activity.[6] This interdependence rules out the possibility of applying a standard combined probability test (Fisher, 1954), and combined tests for dependent p-values require assumptions on the underlying distributions (Kost and McDermott, 2002). It seems that the only sensible way of obtaining a single effect size and p-value is to first average contingency tables and then apply Fisher's test. This is a way to test the statistical significance of the quantity which is ultimately considered, i.e. the effect size averaged over different readout sets.

As a final note to this section, it is worth noticing that the effect size introduced here is slightly different from the one used by Bernardi and Lindner (2017), who assumed that the optimal threshold was learned during the training phase. Consequently, they defined the *maximum* (as

[6] It is possible to get a feeling for these global fluctuations by carefully observing the six realizations of $R^A_\lambda(t)$ in fig. 2.5, which correspond to three different readout sets and two network runs. The six curves can be clearly separated into two bundles corresponding to the two runs, in which each value of λ is present only once. The similar fluctuations displayed by the three curves of different colors in each triplet hint at the presence of global fluctuations correlating the readout activities computed from different neuron sets that belong to the same network.

function of θ_d) of eq. (2.23) as final effect size. This choice raises a technical problem: applying the significance test to the effect size obtained from the same dataset used for the threshold optimization leads to wrong p-values. Hence, the proper procedure has the disadvantage of requiring the generation a dedicated dataset to find the optimal threshold. A detailed discussion on issues related to significance tests and on the different definitions of the effect size is found in appendix B. Therein it is also shown that using a fixed false positive rate, as done here, or an optimized threshold, as done by Bernardi and Lindner (2017), yields mostly similar results.

After presenting how the detection of the stimulus is done in the model, the remainder of this section is concerned with the development of a theory capturing the essential features of the detection process. The purpose of the theory is to obtain analytical estimates of the detection rates and, most importantly, to gain insight into the results of numerical simulations presented in the next section. The first logical step is to characterize theoretically the readout activity, on which the detection is based.

2.3.2. Theoretical characterization of the readout activity

In the asynchronous irregular state, pairwise correlations between spike trains are, on average, weak (Brunel, 2000; Renart et al., 2010). Hence, $R_\lambda^A(t)$ is the filtered sum of a large number of weakly correlated stochastic processes and the central limit theorem applies. If $R_\lambda^A(t)$ can be approximated by a Gaussian process, a complete description of it requires the knowledge of its mean and of its autocorrelation function.

Focusing first on the mean, using eqs. (2.12), (2.13), (2.17) and (2.19) yields:

$$\begin{aligned}
\left\langle \Delta R_\lambda^A(t) \right\rangle &= \left(\frac{1}{C} \sum_{x_i \in \mathcal{S}^A} \langle x_i \rangle \right) \star \mathcal{F}_{\tau_f}(t) - R_{\text{sp}}^A \\
&= \left(\frac{\lambda}{\lambda C} \sum_{x_i \in \mathcal{B}_1} \langle x_i \rangle + \frac{1-\lambda}{(1-\lambda)C} \sum_{x_j \in \mathcal{B}_2} \langle x_j \rangle \right) \star \mathcal{F}_{\tau_f}(t) - r_{\text{sp}} \star \mathcal{F}_{\tau_f}(t) \qquad (2.26) \\
&= (\lambda r_1 + (1-\lambda) r_2) \star \mathcal{F}_{\tau_f}(t) - r_{\text{sp}} \star \mathcal{F}_{\tau_f}(t) \\
&= (\lambda \Delta r_1 + (1-\lambda) \Delta r_2) \star \mathcal{F}_{\tau_f}(t).
\end{aligned}$$

As shown in section 2.2, the time-course of the *unfiltered* firing rates Δr_1 and Δr_2 can be approximated by an exponential relaxation to the new steady-state and back eq. (2.16). Inserting eq. (2.16) into the last equation yields:

$$\left\langle \Delta R_\lambda^A(t) \right\rangle = \Delta r_\lambda \Delta a(t) * \mathcal{F}_{\tau_f}(t), \qquad (2.27)$$

where the height of the step is given by $\Delta r_\lambda = \lambda \Delta r_1 + (1-\lambda)\Delta r_2$. The explicit expression for

the convolution of the two exponential relaxations with the filter is

$$\Delta a(t) * \mathcal{F}_{\tau_f}(t) = \frac{1}{2}\Bigg[\operatorname{erfc}\left(\frac{-2t+3\tau_f}{\sqrt{2}\tau_f}\right) - \exp\left(\frac{\tau_f^2 - 8t\tau_m + 12\tau_f\tau_m}{8\tau_m^2}\right) \operatorname{erfc}\left(\frac{\tau_f^2 - 4t\tau_m + 6\tau_f\tau_m}{2\sqrt{2}\tau_f\tau_m}\right)\Bigg]. \tag{2.28}$$

Equation (2.27) captures well the time-dependent mean activity. To complete the characterization of a Gaussian process, it is necessary and sufficient to specify its second-order statistics, i.e. its autocorrelation function (see section 1.3 p. 19 onwards for definitions of second-order spike-train statistics and spectral measures).

The readout activity cannot be regarded as a stationary process because the stimulation occurs at a specified time. Consequently, the autocorrelation of $R_\lambda^A(t)$ is not a function of a single time argument. To simplify the problem, the following discussion will be restricted to the spontaneous activity (before the stimulus is switched on) by assuming that the effects of the stimulation on the autocorrelation function are negligible. In the stationary state, the knowledge of the autocorrelation function is equivalent to that of its power spectrum, as in eq. (1.14) on p. 20. The power spectrum of the readout activity, $S_{RR}(f)$, can be expressed through the average spike train power spectrum, $S_{xx}(f)$, and the average cross spectrum between spike train pairs, $S_{x_1x_2}(f)$, as the following calculation shows:

$$\begin{aligned}
S_{RR}(f) &= \frac{1}{T}\left\langle \tilde{R}_\lambda^A(f)\tilde{R}_\lambda^{A*}(f)\right\rangle \\
&= \frac{|\tilde{\mathcal{F}}_{\tau_f}(f)|^2}{TC^2} \sum_{x_i,x_j\in\mathcal{S}^A} \langle \tilde{x}_i(f)\tilde{x}_j^*(f)\rangle \\
&= \frac{|\tilde{\mathcal{F}}_{\tau_f}(f)|^2}{TC^2} \sum_{i=j} \langle \tilde{x}_i(f)\tilde{x}_i^*(f)\rangle + \sum_{i<j} \langle \tilde{x}_i(f)\tilde{x}_j^*(f)\rangle + \langle \tilde{x}_j(f)\tilde{x}_i^*(f)\rangle \\
&= |\tilde{\mathcal{F}}_{\tau_f}(f)|^2 \left(\frac{S_{xx}(f)}{C} + \frac{C-1}{C}\Re\left[S_{x_1x_2}(f)\right]\right) \\
&\approx |\tilde{\mathcal{F}}_{\tau_f}(f)|^2 \left(\frac{S_{xx}(f)}{C} + \Re[S_{x_1x_2}(f)]\right),
\end{aligned} \tag{2.29}$$

where $\Re[S_{x_1x_2}(f)]$ is the real part of the average cross-spectrum between neurons. While the power spectrum is on the order of magnitude of the firing rate (i.e. $\approx 2\,\mathrm{Hz}$, see fig. 2.7A), the average cross-spectrum in a sparse network in the AI regime are typically much smaller (see fig. 2.7B). However, the factor $1/C$ renders the contribution of the two terms comparable if C is large, so that none of the two terms can be neglected. The square of the Fourier transformed filter $|\tilde{\mathcal{F}}_{\tau_f}(f)|^2$ decays very rapidly to zero for $f > 10\,\mathrm{Hz}$ (see fig. 2.7, inset). Hence, only the low frequency parts of S_{xx} and $S_{x_1x_2}$ matter to the readout activity.

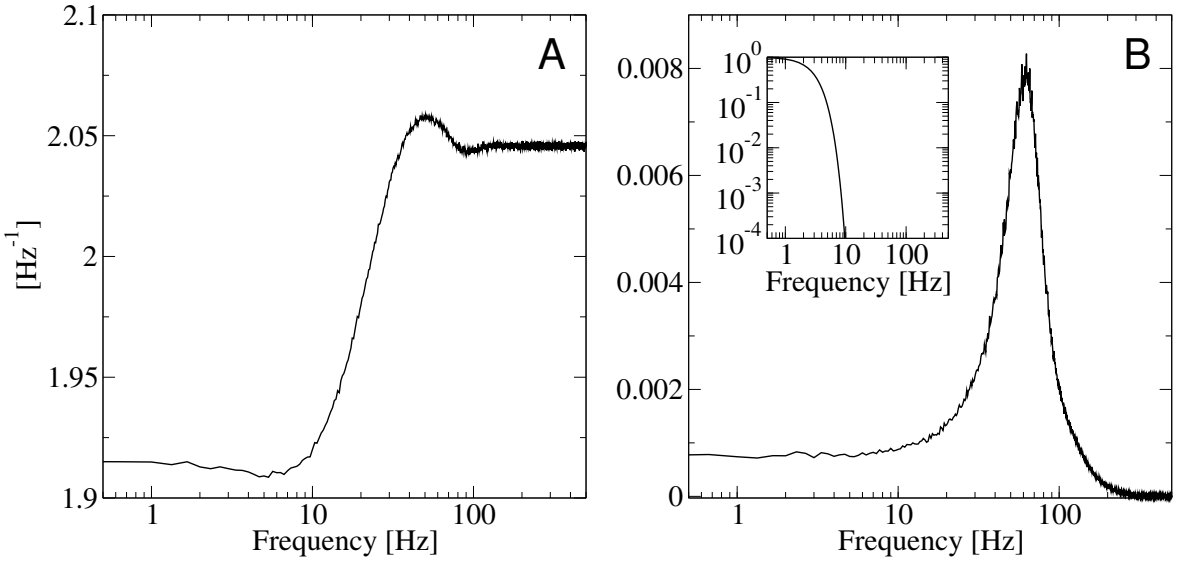

Figure 2.7. – **Spike train power and cross spectrum are almost flat in the low frequency range, in which the filter is substantially larger than zero. A**: single spike train power spectrum $S_{xx}(f)$. **B**: average cross spectrum $S_{x_1x_2}(f)$. The inset shows the rapid decay of $|\tilde{\mathcal{F}}_{\tau_f}(f)|^2$ for $f > 10\,\mathrm{Hz}$. Parameters are here as in table 2.2. Spectra are rather flat in the low frequency range also for the other parameter sets.

As discussed in section 2.1 (fig. 2.2C), the shot-noise theory eq. (2.11) provides a fair approximation to the single-neuron power-spectrum $S_{xx}(f)$. An analytical expression for the other term in eq. (2.29), the average cross-spectrum between spike trains of a recurrent network $S_{x_1x_2}(f)$, is, in general, not straightforward to compute and it is an object of ongoing research (Lindner et al., 2005; Ostojic et al., 2009a; Trousdale et al., 2012; Helias et al., 2013, 2014). The fundamental ansatz of most studies is to treat the recurrent connections as source of a linear perturbation to the dynamics due exclusively to the strong external drive (Lindner et al., 2005). The starting point of these approaches, however, does not apply to the two standard cases of this chapter, in which the external noise is either similar in strength as the recurrent input, or completely absent. The approach pursued in the following does share similarities with the cited examples from the literature (in particular, the linear-response approximation), but does not explicitly require strong *external* noise.

Finding an analytical approximation for $S_{x_1x_2}(f)$ is an interesting problem in itself, because even weak pairwise cross-correlations may have strong effects on the way neuron encode information (Schneidman et al., 2006; Hu et al., 2014). Furthermore, because they are comparatively easy to measure from biological networks, understanding how pairwise correlations between neurons form in a recurrent network model may help to infer the properties of biological networks. However, owing to the focus on the detection problem, in the remainder of this section only the

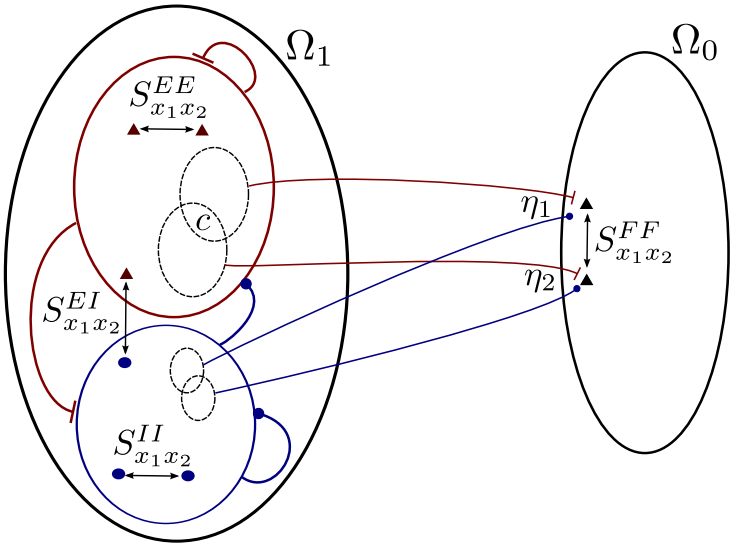

Figure 2.8. – **Auxiliary construction to derive a linear-order approximation to the average cross spectrum.** A second neuronal population Ω_0 receives input from the recurrent network Ω_1 with the same properties as cells inside the recurrent network. However, spikes from Ω_0 have no effect of any other population (including Ω_0 itself). Therefore, $S^{FF}_{x_1x_2}$, the cross spectrum between neurons in Ω_0, is only due to input correlations. The red and blue subset within Ω_1 represent excitatory and inhibitory neurons, respectively. The cross-spectrum between neurons within Ω_1 depends on the type of the considered neuron pair. The three possibilities are $S^{EE}_{x_1x2}$,$S^{EI}_{x_1x2}$,$S^{II}_{x_1x2}$ for excitatory-excitatory, excitatory-inhibitory, inhibitory-inhibitory.

main ideas of the derivation are outlined, while the details of the rather lengthy full calculation and the comparisons with numerical simulations are reported in appendix A.

As a starting point, an auxiliary construction is needed: imagine a set of neurons receiving the same kind of input from the recurrent network as any neuron within the network, but lacking any output connection. For brevity, this new set of neurons is indicated from now on as Ω_0 and the recurrent network as Ω_1 (see fig. 2.8). More precisely, each neuron in Ω_0 receives spike trains from C_E excitatory and C_I inhibitory neurons selected independently at random within Ω_1 with connection weights and delays drawn from the same distributions as for the recurrent weights. As pairs of neurons in Ω_0 are not connected to each other, cross-correlations between them can only be due to cross-correlations between their inputs η_1 and η_2. If these input cross-correlations are not too strong, the cross-spectrum between two neurons in Ω_0, $S^{FF}_{x_1x_2}(f)$, is approximately given by the product of the input cross-spectrum $S_{\eta_1\eta_2}(f)$ with the absolute square of $\chi(f)$, the firing-rate susceptibility to a weak perturbation of the input current (details in appendix A, see also Ostojic et al., 2009a):

$$S^{FF}_{x_1x_2}(f) = |\chi(f)|^2 \frac{\langle \tilde{\eta}_1 \tilde{\eta}_2^* \rangle}{T} = |\chi(f)|^2 S_{\eta_1\eta_2}(f). \tag{2.30}$$

Even if the background network-noise is approximated as white shot-noise, the susceptibility with respect to additive input is not known.[7] However, the so-called DC susceptibility, i.e. the firing-rate response for a slow-varying signal $\chi(f \to 0)$, can be obtained by taking the derivative of the stationary firing rate eq. (2.9) with respect to the mean input $\mu = R_m I_0$ (see section 1.3 p. 30). A rough estimate of $\chi(f)$ for frequencies larger than zero can be obtained by recalling that an exponential decay with time constant τ_m provided a fair approximation to the time course of the time-dependent firing rate after stimulation. Translating this approximation into the frequency domain gives:

$$\chi(f) \approx \frac{\chi(0)}{1 - 2\pi i f \tau_m} = \frac{\mathrm{d}\phi_{sn}/\mathrm{d}\mu}{1 - 2\pi i f \tau_m}. \tag{2.31}$$

Because in the remainder of this section only quantities in the Fourier representation appear, to simplifiy the notation all tildes and frequency arguments of Fourier-transformed spike trains and of spectra will be omitted in the following. The average cross-spectrum between the two input currents $S_{\eta_1\eta_2}$ can be expressed through the single-neuron average power-spectrum S_{xx} and of the three cross-spectra $S^{EE}_{x_1x_2}$, $S^{II}_{x_1x_2}$, and $S^{EI}_{x_1x_2}$, which are the average cross-spectra between two excitatory neurons, between two inhibitory neurons, and between an excitatory and an inhibitory neuron in the recurrent network, respectively (see fig. 2.8). The calculation is straightforward but lengthy, and is reported in appendix A. The result is:

$$\begin{aligned} S_{\eta_1\eta_2} = \tau_m^2 J^2 |\mathcal{D}(f)|^2 \{ & p_c C_E (1 + g^2\gamma) S_{xx} + S^{EE}_{x_1x_2}(C_E^2 - p_c C_E) \\ & - 2g\gamma C_E^2 \Re[S^{EI}_{x_1x_2}] + g^2 S^{II}_{x_1x_2}(\gamma^2 C_E^2 - p_c \gamma C_E)\}. \end{aligned} \tag{2.32}$$

In the last equation, $\mathcal{D}(f)$ is the characteristic function of the delay distribution

$$|\mathcal{D}(f)|^2 = \frac{2 - 2\cos(2\pi f \Delta)}{(2\pi f \Delta)^2} \tag{2.33}$$

where $\Delta = D_{\max} - D_{\min}$. Differences in arrival times due to different delays disrupt input cross-correlations, but shifting all delays by a fixed quantity would not produce any effect. Therefore, the spread of the delay distribution Δ determines how fast input cross-correlations fade for increasing frequency, but the average delay does not appear in eq. (2.33).

By construction, neurons in Ω_0 and Ω_1 receive statistically equivalent input, so that their output should be the same. However, there is a difference when pairs of neurons are considered:[8] neurons in Ω_0 cannot influence each other, whereas the output spikes fired by one of the two

[7] The shot-noise theory by Richardson and Swarbrick (2010) includes the linear response to a variation of the input *rates* but not for an additive input. Droste and Lindner (2017a) calculated the susceptibility to an additive signal for exponential excitatory shot-noise background but without inhibitory shot-noise input.

[8] In principle, a similar difference exists for single neurons: a neuron in Ω_1 can affect itself via loops of length two or longer. However, the effect on single-neuron statistics, i.e. firing rate and the power spectrum, is negligible.

neurons in Ω_1 can reach and affect the other one, either directly (if the two neurons happen to be connected with each other) or via longer paths. Considering two neurons in Ω_1, the following linear-response ansatz for the activity of the second neuron (here $X, Y = E, I$ indicate the type of neuron one and two, respectively) can be made:

$$x_2^Y \approx x_{2,0} + \sum_{\ell=1}^{\infty} \mathcal{L}_{\ell,1}^X x_1^X, \tag{2.34}$$

where $x_{2,0}$ is the activity of the second neuron if the effect of the spikes fired from neuron one is removed from the network, and $\mathcal{L}_{\ell,1}^X$ summarizes the effect of spikes fired by neuron one reaching neuron two via paths of length ℓ. Because the network is large, the spikes fired from neuron one are a small fraction of the total spikes fired by the network, which justifies the linear-response assumption. Because the labels for the two neurons are arbitrary, the indexes in eq. (2.34) can be swapped

$$x_1^X \approx x_{1,0} + \sum_{\ell=1}^{\infty} \mathcal{L}_{\ell,2}^Y x_2^Y. \tag{2.35}$$

Equations (2.34) and (2.35) can be combined and solved for x_1, x_2, which, under some assumptions discussed in appendix A, leads to

$$\begin{aligned} S_{x_1x_2}^{XY} &= \frac{1}{T}\langle x_1 x_2^* \rangle \approx \frac{1}{T}\langle x_{1,0} x_{2,0}^* \rangle + \left[\left\langle (\sum_{\ell=1}^{\infty} \mathcal{L}_{\ell,1}^X)^* \right\rangle + \left\langle \sum_{\ell=1}^{\infty} \mathcal{L}_{\ell,2}^Y \right\rangle \right] S_{xx}, \\ &= S_{x_1x_2}^{FF} + \left[\left(\sum_{\ell=1}^{\infty} \mathcal{L}_{\ell}^X \right)^* + \sum_{\ell=1}^{\infty} \mathcal{L}_{\ell}^Y \right] S_{xx}, \end{aligned} \tag{2.36}$$

where $\mathcal{L}_{\ell}^X = \langle \mathcal{L}_{\ell,1}^X \rangle$ and $\mathcal{L}_{\ell}^Y = \langle \mathcal{L}_{\ell,2}^Y \rangle$. In eq. (2.36), the identification $\frac{1}{T}\langle x_{1,0} x_{2,0}^* \rangle = S_{x_1x_2}^{FF}$ is possible because $x_{1,0}$ and $x_{2,0}$ indicate the activity of the two neurons if the effect of all spikes fired by the two considered neurons themselves via the recurrent network is removed, which makes them fully equivalent to neurons in Ω_0.

As an example, the contribution of direct connections can be first isolated and then calculated explicitly. A simple rearrangement of eq. (2.36) yields

$$\begin{aligned} S_{x_1x_2}^{XY} &= S_{x_1x_2}^{FF} + \left[(\mathcal{L}_1^X)^* + \mathcal{L}_1^X \right] S_{xx} + \left[(\sum_{\ell=2}^{\infty} \mathcal{L}_{\ell}^X)^* + \sum_{\ell=2}^{\infty} \mathcal{L}_{\ell}^Y \right] S_{xx} \\ &= S_{x_1x_2,nc}^{XY} + \left[(\mathcal{L}_1^X)^* + \mathcal{L}_1^Y \right] S_{xx}, \end{aligned} \tag{2.37}$$

where $S_{x_1x_2,nc}^{XY}$ indicates the cross-spectrum between neurons in Ω_1 that are not directly connected to each other. When dealing with direct connections, there are three plus one possible motifs: i) the first neuron is connected to the second one; ii) the second neuron is connected to the first one; iii) they are mutually connected; or iv) the two neurons are not directly connected.

For instance, if the two neurons are excitatory and the first one is connected to the second one, then

$$x_2^E \approx x_{2,0} + \chi\tau_m J_{21} e^{i2\pi f D_{21}} x_1^E, \tag{2.38}$$

where $x_{2,0}$ indicates here - with a slight abuse of notation - the activity of neuron two if the direct connection is removed (but including the effect of longer paths). Using the susceptibility realization-wise as in eq. (2.38) was put forward by Lindner et al. (2005). The last equation can be used to express the cross-spectrum of excitatory pairs in which the first neuron is connected to the second as:

$$S^{EE}_{x_1x_2,1\to2} = \frac{1}{T}\langle x_1 x_2^*\rangle \approx \frac{1}{T}\langle x_1 x_{2,0}^*\rangle + \frac{1}{T}\chi^*\tau_m J\mathcal{D}^*(f)\langle x_1 x_1^*\rangle = S^{EE}_{x_1x_2,nc} + \chi^*\tau_m J\mathcal{D}^*(f)S_{xx}. \tag{2.39}$$

If the same linear-response ansatz is applied to each motif, similar expressions can be obtained for each of the four possible cases. Summing the results for the four cases, each multiplied by the probability for the corresponding motif to occur, yields

$$S^{EE}_{x_1x_2} = S^{EE}_{x_1x_2,nc} + \tau_m J 2\Re[\chi(f)\mathcal{D}(f)]p_c S_{xx}. \tag{2.40}$$

From the comparison of the last equation with eq. (2.37), one deduces that $\mathcal{L}_1^E = p_c A$ where

$$A(f) = \tau_m J\chi(f)\mathcal{D}(f), \tag{2.41}$$

summarizes the average linear response of a neuron's firing rate to a single excitatory spike train. Similarly, one finds from analogous expressions for the EI- and II-pairs that $\mathcal{L}_1^I = -g\mathcal{L}_1^E$.

Although the effect of a single path of length ℓ decreases with ℓ, the number of possible paths connecting two neurons increases with ℓ, so that their contribution is not negligible. The combined effect of paths of all lengths is examined in appendix A, and results in

$$\begin{aligned}\sum_{\ell=1}^{\infty}\mathcal{L}_\ell^E &= \frac{1}{1 - Ap_cN_E(1-g\gamma)} = \beta_{\mathcal{L}} \\ \sum_{\ell=1}^{\infty}\mathcal{L}_\ell^I &= \frac{-g}{1 - Ap_cN_E(1-g\gamma)} = -g\beta_{\mathcal{L}}.\end{aligned} \tag{2.42}$$

If eq. (2.42) is combined with eq. (2.36), one finds

$$S^{EE}_{x_1x_2} \approx S^{FF}_{x_1x_2} + p_c 2\Re[A\beta_{\mathcal{L}}]S_{xx} \tag{2.43}$$

$$S^{EI}_{x_1x_2} \approx S^{FF}_{x_1x_2} + p_c\left[(A\beta_{\mathcal{L}})^* - gA\beta_{\mathcal{L}}\right]S_{xx} \tag{2.44}$$

$$S^{II}_{x_1x_2} \approx S^{FF}_{x_1x_2} - p_c g 2\Re[A\beta_{\mathcal{L}}]S_{xx}. \tag{2.45}$$

Equations (2.43) to (2.45) can be inserted into eqs. (2.30) and (2.32). From this substitution

an equation containing only $S^{FF}_{x_1x_2}$ and S_{xx} is obtained. Solving for $S^{FF}_{x_1x_2}$, neglecting terms of order $\sim 1/N_E$, and rearranging yields

$$S^{FF}_{x_1x_2} \approx |A|^2 p_c C_E (1+g^2\gamma)|\beta_{\mathcal{L}}|^2 S_{xx} = \frac{|A|^2 p_c C_E (1+g^2\gamma)}{|1-AC_E(1-g\gamma)|^2} S_{xx}. \tag{2.46}$$

Equation (2.46) can be inserted into eqs. (2.43) to (2.45). The final result is the average of $S^{EE}_{x_1x_2}$, $S^{EI}_{x_1x_2}$, and $S^{II}_{x_1x_2}$ with weights corresponding to the respective number of EE, EI, IE, and II pairs, keeping in mind that $S^{IE}_{x_1x_2} = (S^{EI}_{x_1x_2})^*$ for symmetry. With the further approximation $N(N-1) \approx N^2$, the final result is

$$\begin{aligned} S_{x_1x_2} &= \left(\frac{N_E}{N_E(1+\gamma)}\right)^2 S^{EE}_{x_1x_2} + \gamma \left(\frac{N_E}{N_E(1+\gamma)}\right)^2 (S^{EI}_{x_1x_2} + S^{IE}_{x_1x_2}) + \left(\frac{\gamma N_E}{N_E(1+\gamma)}\right)^2 S^{II}_{x_1x_2} \\ &= S^{FF}_{x_1x_2} + \frac{2p_c \Re[A\beta_{\mathcal{L}}]}{1+\gamma}(1-g\gamma)S_{xx} \\ &= S_{xx} p_c \left(|A|^2 \frac{C_E(1+g^2\gamma)}{|1-AC_E(1-g\gamma)|^2} + 2\Re\left[\frac{A}{1-AC_E(1-g\gamma)}\right]\frac{1-g\gamma}{1+\gamma}\right). \end{aligned} \tag{2.47}$$

The comparison of eq. (2.47) with simulation results is discussed in detail in appendix A and shown in fig. A.9 on p. 223. For the standard parameters both in the presence and in the absence of external shot-noise (tables 2.1 and 2.2), the theory underestimates the cross-spectrum by about 50% in the low-frequency range, which is the relevant one as far as the spectrum of readout activity is concerned (see fig. 2.7). For the "single-barrel" parameter set introduced in section 2.6 (table 2.3), however, the agreement at low-frequencies is reasonable. Therefore, only in this case eq. (2.47) will be actually used with eq. (2.29) to approximate the power-spectrum of the readout activity. However, eq. (2.47) will also prove useful in section 2.5 when the two standard parameter sets are used to interpret the qualitative behavior of the network fluctuations when the network size is scaled, despite the quantitative discrepancies with simulations.

Lastly, it is worth noting that eq. (2.47) is the same as eq. (25) in the paper by Trousdale et al. (2012). On the one hand, it may be expected because both approaches consider expansions to the linear order and sum contributions from different correlation sources. On the other hand, it is somewhat surprising that they turn out to be *completely* equivalent, because Trousdale et al. (2012) assume strong external drive, whereas the approach followed here does not. A possible solution to the paradox is perhaps that the uncorrelated part of the network input to each neuron plays the role of the external noise (a slightly more detailed discussion about similarities and differences in the two calculations is found in appendix A).

2.3.3. Theoretical approximation of detection rates

In the last subsection, the readout activity has been characterized as a Gaussian process and analytical approximations for its mean and autocorrelation function have been discussed. The next step is, assuming the knowledge of $\left\langle R_\lambda^A(t)\right\rangle$ and $S_{RR}(f)$, to estimate the detection rates and the effect size.

Correct detection and false positive rates are the fraction of trajectories that, starting from stationary initial conditions, escape *at least once* from the area enclosed by two detection barriers within a time T_w in the presence and in the absence of the stimulus, respectively. Hence, the two detection rates are related to an integral over the first-passage-time density of the non-stationary stochastic process $R_\lambda^A(t)$. The exact first-passage-time density for a Gaussian process is only known for several special cases, and finding a solution for a generic autocorrelation function is a hard problem. In the following, instead of pursuing the first-passage-time density, a drastic approximation of the escape problem will be used to obtain a fair estimate of the detection rates.

The basic idea is to replace the probability for the Gaussian *process* to leave the interval $\left(R_{\mathrm{sp}}^A-\theta_d, R_{\mathrm{sp}}^A+\theta_d\right)$ in the time window T_w with the probability for at least one out of n independent draws of a Gaussian *variable* to fall outside the same interval, as schematically portrayed in fig. 2.9. To obtain the number of draws suitable to represent the continuous-time problem, one can start considering the probability of finding the readout activity within the two boundaries at a given time t. For a Gaussian process, this probability is given by the integral

$$\begin{aligned} p_{\Delta R}\left(\theta_d, \Delta R_\lambda^A(t)\right) &= \int_{-\theta_d}^{\theta_d} \mathrm{d}x\, \mathcal{N}\left(x, \Delta R_\lambda^A(t), \sigma_A^2\right) \\ &= \frac{1}{2}\left[\mathrm{erf}\left(\frac{\theta_d - \left\langle \Delta R_\lambda^A(t)\right\rangle}{\sqrt{2\sigma_A^2}}\right) + \mathrm{erf}\left(\frac{\theta_d + \left\langle \Delta R_\lambda^A(t)\right\rangle}{\sqrt{2\sigma_A^2}}\right)\right], \end{aligned} \tag{2.48}$$

where σ_A^2, the variance of $R_\lambda^A(t)$, has been treated as independent of time[9] and $\Delta R_\lambda^A(t) = R_\lambda^A(t) - R_{\mathrm{sp}}^A$. Equation (2.48) yields the probability for the outcome of the first draw. Shortly thereafter, the value of the activity will be highly correlated with its value at t, so that the probability that the readout activity is again found outside the boundaries strongly depends on the outcome of the first draw. However, after a certain amount of time, past values are forgotten and the probability of finding ΔR_λ^A within the two thresholds becomes independent of the first draw and eq. (2.48) can be used again. An estimate for the time necessary for the process to

[9] As already stated on p. 57, the present analysis ignores the effect of the stimulus on the second-order statistics of $R_\lambda^A(t)$. With this assumption, variance and autocorrelation function of $R_\lambda^A(t)$ do not depend on time and can be obtained from the power spectrum $S_{RR}(f)$ (see section 1.3).

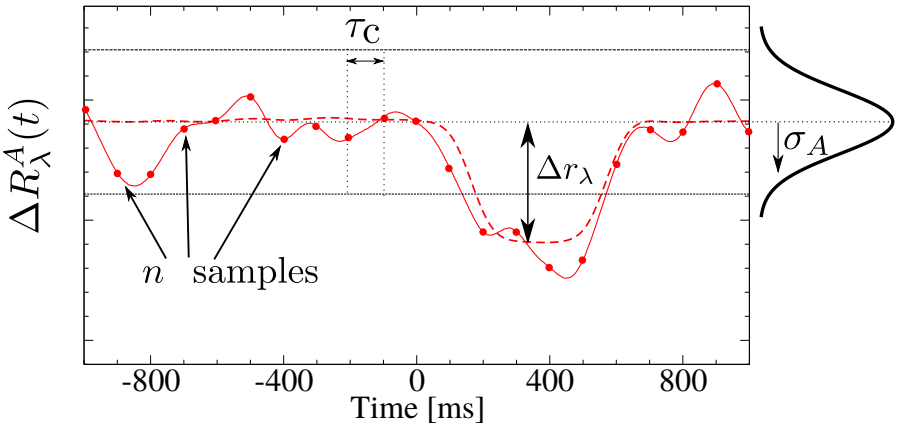

Figure 2.9. – **Discrete approximation to first-passage time of the readout activity.**

forget its past, i.e. interval between draws, is the correlation time

$$\tau_c = \int_0^\infty \mathrm{d}t \frac{|C_{RR}(t)|}{C_{RR}(0)}, \tag{2.49}$$

where $C_{RR}(t)$ is the autocorrelation function of $R^A_\lambda(t)$. From these considerations it follows that the number of independent draws can be taken as $n = T_w/\tau_c$.

For $t < 0$, the process is stationary because the stimulus is absent. Therefore, the probability of each draw does not depend on t and eq. (2.48) reduces to

$$p_0(\theta_d) = p_{\Delta R}\left(\theta_d, \Delta R^A_\lambda(t<0)\right) = p_{\Delta R}\left(\theta_d, 0\right) = \mathrm{erf}\left(\frac{\theta_d}{\sqrt{2}\sigma_A}\right). \tag{2.50}$$

In this discretized description, the false positive rate is the probability that *at least one* draw falls outside $(-\theta_d, \theta_d)$. The probability for the *opposite* case, namely that all draws fall within the two barriers, is simply $p_0(\theta_d)^n$. Hence, the false positive rate is

$$\mathcal{FP}(\theta_d) \approx 1 - p_0^n(\theta_d). \tag{2.51}$$

The correct detection rate can be estimated similarly, with the difference that the probability of each draw here depends on t:

$$\mathcal{CD}(\theta_d) \approx 1 - \prod_{k=1}^{n} p_{\Delta R}\left(\theta_d, \Delta R^A_{\lambda,k}\right), \tag{2.52}$$

where $\Delta R^A_{\lambda,k}$ are n values suitably chosen to represent the time course of $\Delta R^A_\lambda(t)$ in the correct detection window $(0, T_w)$. One possibility is to take the mean value of $\Delta R^A_\lambda(t)$ in each segment

$(k\tau_c, k\tau_c + \tau_c)$

$$\overline{\Delta R}^A_{\lambda,k} = \frac{1}{\tau_c}\int_{k\tau_c}^{(k+1)\tau_c} \mathrm{d}t\, \Delta R^A_\lambda(t), \tag{2.53}$$

which leads to

$$\mathcal{CD}_{\mathrm{avg}}(\theta_d) \approx 1 - \prod_{k=1}^{n} p_{\Delta R}\left(\theta_d, \overline{\Delta R}^A_{\lambda,k}\right). \tag{2.54}$$

A more conservative choice is to take the minimum absolute deviation in each segment

$$\widehat{\Delta R}^A_{\lambda,k} = \Delta R^A_\lambda\left(\underset{k\tau_c<t<(k+1)\tau_c}{\arg\min}\left\{\left|\Delta R^A_\lambda(t)\right|\right\}\right) \tag{2.55}$$

to obtain a "lower bound" estimate for the correct detection rate

$$\mathcal{CD}_{\mathrm{LB}}(\theta_d) \approx 1 - \prod_{k=1}^{n} p_{\Delta R}\left(\theta_d, \widehat{\Delta R}^A_{\lambda,k}\right). \tag{2.56}$$

In both cases, the effect size as a function of θ_d can be found by subtracting eq. (2.51) from eq. (2.54) or eq. (2.56), i.e.

$$\mathcal{Y}_{\mathrm{avg}}(\theta_d) = p_0^n(\theta_d) - \prod_{k=1}^{n} p_{\Delta R}\left(\theta_d, \overline{\Delta R}^A_{\lambda,k}\right) \tag{2.57}$$

$$\mathcal{Y}_{\mathrm{LB}}(\theta_d) = p_0^n(\theta_d) - \prod_{k=1}^{n} p_{\Delta R}\left(\theta_d, \widehat{\Delta R}^A_{\lambda,k}\right). \tag{2.58}$$

Any of the two expressions eqs. (2.57) and (2.58) can be used to provide an analytical prediction of the simulation results. However, both equations are not very legible and, hence, unfit to provide insights into the qualitative behavior of numerical results. Progress can be made in this direction by neglecting the actual time course of ΔR^A_λ and replacing it with a box-shaped function of length T_s and height Δr_λ. With this simplification, $n_s = T_s/\tau_c$ independent draws with the same probability of being within the two barriers are obtained. This probability, denoted by p_1, is found by setting $\left\langle \Delta R^A_\lambda(t)\right\rangle = \Delta r_\lambda$ in eq. (2.48). By doing so and introducing the *(signed) signal-to-noise ratio*

$$\delta^A_\lambda = \frac{\Delta r_\lambda}{\sigma_A} \tag{2.59}$$

one obtains

$$p_1(\theta_d, \delta^A_\lambda) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{\theta_d}{\sqrt{2}\sigma_A} + \frac{\delta^A_\lambda}{\sqrt{2}}\right) + \mathrm{erf}\left(\frac{\theta_d}{\sqrt{2}\sigma_A} - \frac{\delta^A_\lambda}{\sqrt{2}}\right)\right]. \tag{2.60}$$

For the remaining $n - n_s$ draws, the probability is $p_0(\theta_d)$ as for the false positive rate. Hence,

the correct detection rate reduces to

$$\mathcal{CD}_{\text{box}}(\theta_d) \approx 1 - p_1^{n_s}(\theta_d, \delta_\lambda^A) p_0^{n-n_s}(\theta_d). \tag{2.61}$$

The last expression can be used in combination with eq. (2.51) to derive an explicit expression for the the hit rate as a function of the false positive rate, i.e. for the ROC curve (see p. 54). To this end, one can solve eq. (2.51) for θ_d and insert it into eq. (2.61):

$$\mathcal{CD}_{\text{box}}(\mathcal{FP}) \approx 1 - \left(\frac{1}{2}\left[\text{erf}\left(\mathcal{E}_n(\mathcal{FP}) + \frac{\delta_\lambda^A}{\sqrt{2}}\right) + \text{erf}\left(\mathcal{E}_n(\mathcal{FP}) - \frac{\delta_\lambda^A}{\sqrt{2}}\right)\right]\right)^{n_s} (1 - \mathcal{FP})^{\frac{n-n_s}{n}} \tag{2.62}$$

where the auxiliary function $\mathcal{E}_n(x) = \text{erf}^{-1}((1-x)^{\frac{1}{n}})$ was introduced. Setting $\mathcal{FP} = 0.25$ in the last equation yields the effect size:

$$\bar{\mathcal{Y}}(\lambda) = \frac{3}{4} - \left(\frac{1}{2}\left[\text{erf}\left(\mathcal{E}_n\left(\frac{1}{4}\right) + \frac{\delta_\lambda^A}{\sqrt{2}}\right) + \text{erf}\left(\mathcal{E}_n\left(\frac{1}{4}\right) - \frac{\delta_\lambda^A}{\sqrt{2}}\right)\right]\right)^{n_s} \left(\frac{3}{4}\right)^{\frac{n-n_s}{n}}. \tag{2.63}$$

2.3.4. Effect size and signal-to-noise ratio

Equations (2.62) and (2.63) reveal that the ROC and the effect size are completely determined (in the simplified Gaussian approximation) by the parameters n_s, n, and by the signal-to-noise ratio δ_λ^A. As the filter time scale is considerably larger than the unfiltered network activity, it turns out that $\tau_c \approx \tau_f$. Because the effect size does not depend crucially on the exact value of τ_c, it is convenient to perform the further simplification $\tau_c = \tau_f$. If the detection time window T_w and the stimulus duration T_s are fixed (in the experiment they were not varied), the effect size depends uniquely on δ_λ^A.

The effect size as a function of the signal-to-noise ratio (SNR) is shown in fig. 2.10. The theory is eq. (2.63) with the simplification $\tau_c = \tau_f$, while data points for excitatory (black circles) and inhibitory (red squares) $\mathcal{B}_0$ are for the case with external shot noise. The effect size is an even function and it is a monotonic increasing function of $|\delta_\lambda^A|$. Figure 2.10 shows that the theory tends to overestimate the effect size, especially for large SNRs and when $\mathcal{B}_0$ is inhibitory. Possible reasons for this are discussed in the following section.

To establish a link between the effect size and the properties of the system, it is useful to reexamine the definition of the SNR and to further approximate it. Starting with the numerator, one obtains:

$$\delta_\lambda^A = \frac{\Delta r_\lambda}{\sigma_A} = \frac{\lambda \Delta r_1 + (1-\lambda)\Delta r_2}{\sigma_A} \approx \frac{\lambda \Delta r_1}{\sigma_A} \approx \frac{\lambda \tau_m J^X \chi(0) \Delta r_0}{\sigma_A}, \tag{2.64}$$

where the first approximation is justified by the fact that $\Delta r_1 \gg \Delta r_2$ and is valid only for not too small values of λ. The second approximation in eq. (2.64) neglects the contribution of the

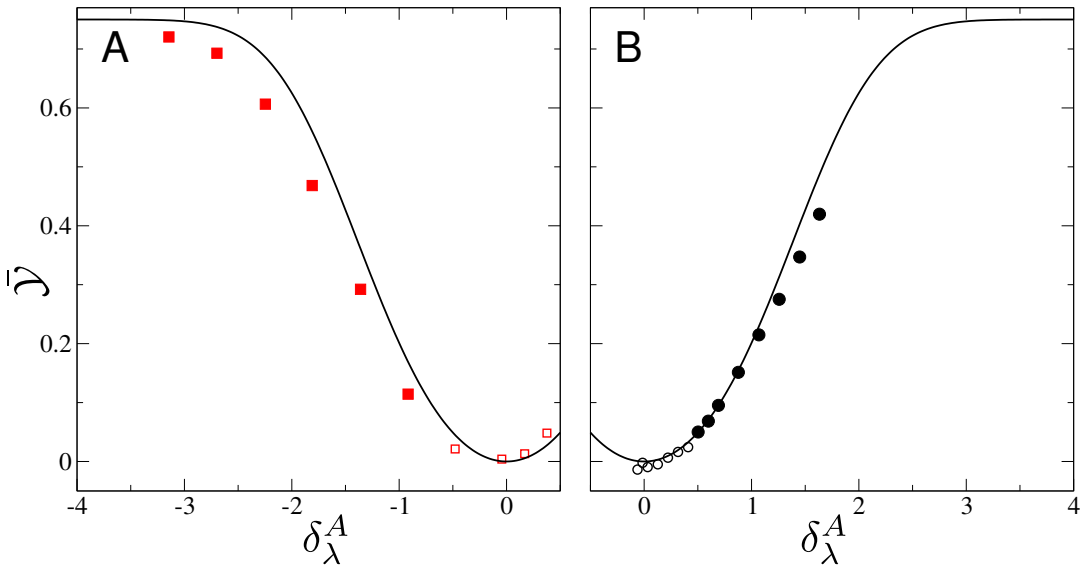

Figure 2.10. – **Functional relationship between effect size and signal-to-noise ratio.** The effect size $\bar{\mathcal{Y}}$ is plotted as a function of the signal-to-noise ratio δ_λ^A. The theoretical line is eq. (2.63). Data points are simulation results with external shot noise (parameters as in table 2.2) for inhibitory (**A**) and excitatory (**B**) $\mathcal{B}_0$. Filled symbols are data points that are significantly different from zero (p-value smaller than 0.05). The effect size is a monotonically increasing function of $|\delta_\lambda^A|$. The signal-to-noise ratio cannot be directly prescribed in simulations, and was indirectly controlled by varying λ.

recurrent connections to Δr_1, the change in the firing rate of the cells in $\mathcal{B}_1$, and introduces the DC susceptibility $\chi(0)$, already used in eq. (2.31). The symbol J^X indicates the average output weight of $\mathcal{B}_0$, which depends on the cell type $X = E, I$.

The denominator of the SNR is σ_A, the standard deviation of the readout activity $R_\lambda^A(t)$. By first using eq. (1.15) (p. 20) to express the variance σ_A^2 as the integral over the power spectrum $S_{RR}(f)$ and then exploiting eq. (2.29), one can split σ_A^2 into two parts, the first one proportional to the low-frequency limit of the single spike-train power-spectrum $S_{xx}(0)$ and the other one proportional to the low-frequency limit of the cross-spectrum between neurons $S_{x_1x_2}(0)$:

$$\begin{aligned}\sigma_A^2 &= \int_{-\infty}^{+\infty} \mathrm{d}f\, S_{RR}(f) = \int_{-\infty}^{+\infty} \mathrm{d}f |\tilde{\mathcal{F}}_{\tau_f}(f)|^2 \left(\frac{S_{xx}(f)}{C} + S_{x_1x_2}(f) \right) \\ &\approx \left(\frac{S_{xx}(0)}{C} + S_{x_1x_2}(0) \right) \int_{-\infty}^{+\infty} \mathrm{d}f\, |\tilde{\mathcal{F}}_{\tau_f}(f)|^2 \\ &\approx \frac{S_{xx}(0)}{C\sqrt{\pi}\tau_f} + \frac{S_{x_1x_2}(0)}{\sqrt{\pi}\tau_f} = \frac{s_{xx}}{C} + s_{x_1x_2}\end{aligned} \tag{2.65}$$

The first approximation in eq. (2.65) hinges upon the fact that power and cross-spectra are rather flat for frequencies up to $1/\tau_f = 10\,\mathrm{Hz}$, above which the filter decays very rapidly to zero so that higher frequencies do not contribute to the integral (see fig. 2.7). The second approximation neglects the two Heaviside functions in the definition of the filter.

In the end, the SNR can be approximated as

$$\delta_\lambda^A = \frac{\Delta r_\lambda}{\sigma_A} \approx \lambda \tau_m J^X \frac{\chi(0)\Delta r_0}{(s_{xx}/C + s_{x_1x_2})^{1/2}}. \tag{2.66}$$

This expression will be useful in the next sections to interpret the result of network simulations. One caveat to keep in mind is, however, that several of the quantities appearing in eq. (2.66) ($\chi(0)$, Δr_0, s_{xx}, and $s_{x_1x_2}$) are not model parameters with a prescribed value but result from the network dynamics. Therefore, they depend implicitly - and often non-trivially - on the network parameters and on each other.

2.4. Detectability of single-cell stimulation

The previous section has laid the prerequisites to answer the main question of this chapter by introducing a rather simple detection procedure that considers $R_\lambda^A(t)$, the activity of a readout subset $\mathcal{S}^A$, and reacts to deviations from the spontaneous value. The most important parameter of the detector is λ, which quantifies the bias when selecting the neurons forming the readout set. The detector's response is either classified as a hit (a correct detection), when the stimulus was actually present, or as a false positive, if the stimulus was not present. The probability of hits and false positives is influenced by the detector's sensitivity, i.e. by the value of the decision threshold θ_d. The final output of the model, the *effect size* $\bar{\mathcal{Y}}$, is defined as the difference between correct detection and false positive rate when θ_d is set to obtain a false positive rate of 25%, which corresponds approximately to the false positive rate measured in the experiments.

Effect size as a function of the bias

The case of the autonomous network, i.e. receiving no external noise, will be considered first. The central quantity to describe the detectability of the stimulation, the effect size $\bar{\mathcal{Y}}$, is shown in fig. 2.11A as a function of the bias of the detector λ. Data points that are significantly different than zero (p-value < 0.05), are represented by filled up symbols. When the readout is not biased and λ is left at its natural value $\lambda = \lambda_0 = C/N = 0.05$, the stimulation is not detectable, regardless of the type of the stimulated cell. Because λ represents the effects of the training phase, this result agrees with the experimental finding that the stimulation is not detectable in the untrained system. However, if the value of λ is increased, the effect size grows. The increment is very rapid when considering an inhibitory $\mathcal{B}_0$ (red squares): the first statistically significant effect data point (for $\lambda = 0.15$) marks a rather large effect size of $\bar{\mathcal{Y}} \approx 15\%$, and for $\lambda \geq 0.4$ the effect saturates in the vicinity of the maximal effect 75%. If the stimulated cell is excitatory (black circles), the effect size increases more slowly. Still, a bias of $\lambda = 0.2$ suffices to obtain statistically significant detection with an effect size of about 5%. The theory (continuous lines of corresponding color) was obtained by using eq. (2.63), in which the value of the SNR δ_λ^A is semi-analytical: its numerator Δr_λ was computed from eqs. (2.14) and (2.15), while its denominator σ_A is measured from simulations (see discussion in section 2.3.2 and appendix A). Considering the many approximations involved in the derivation of the theory, its agreement with the simulations is quite satisfactory. As already observed in fig. 2.10, the strongest discrepancies are observed for strong bias signals and when $\mathcal{B}_0$ is inhibitory.

Adding external shot-noise to the network makes only a small quantitative difference, as fig. 2.11B shows (the meaning of all symbols and colors is the same as in the previous case). The effect size for a given λ is here in general slightly smaller than in the absence of external noise. Conversely, the necessary bias to achieve a statistically significant detection is larger.

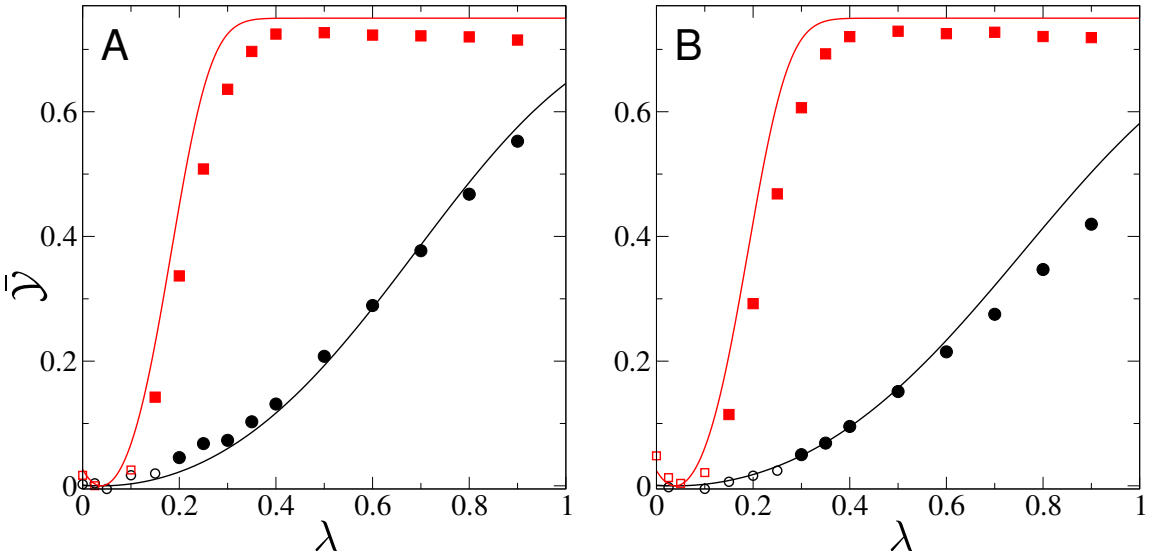

Figure 2.11. – **The single cell stimulation is detectable for moderate values of the bias both in the presence and in the absence of external shot-noise.** Effect size $\bar{\mathcal{Y}}$ as a function of the bias parameter λ in the absence (**A**) and presence (**B**) of external shot-noise input. Black circles (red squares) indicate simulation results for the case of excitatory (inhibitory) $\mathcal{B}_0$. Closed symbols indicate data points that are significantly different than zero (significance level $p < 0.05$). Theoretical lines are based on eq. (2.63). The value of σ_A used to compute eq. (2.63) was measured from stimulations. Parameters in **A** as in table 2.1 with $N_{\text{trials}} = 800$ for excitatory $\mathcal{B}_0$ and $N_{\text{trials}} = 400$ for inhibitory $\mathcal{B}_0$. Parameters in **B** as in table 2.2 with $N_{\text{trials}} = 900$ for excitatory $\mathcal{B}_0$ and $N_{\text{trials}} = 400$ for inhibitory $\mathcal{B}_0$.

However, the qualitative picture is unchanged: in the untrained system (for $\lambda = \lambda_0$), the single-cell stimulation is not detectable. However, if the readout is biased (a caricature for the training phase in the experiments), the stimulation is detectable. For a given amount of learning, stimulating an inhibitory cell produces a larger effect.

Quite remarkably, all these results are consistent with the experiments: the naive animals cannot report the stimulation, and after the training period inhibitory cells are more easily detectable than excitatory ones. The better detectability of inhibitory $\mathcal{B}_0$ in the model can be easily explained by the fact that inhibitory weights are on average stronger than excitatory ones. More precisely, in the approximation eq. (2.64), the SNR is proportional to J^X,

$$\delta_\lambda^A = \frac{\lambda\Delta r_1 + (1-\lambda)\Delta r_2}{\sigma_A} \approx \frac{\lambda\tau_m J^X \chi(0)\Delta r_0}{(s_{xx}/C + s_{x_1x_2})^{1/2}}, \tag{2.67}$$

where $J^E = J$ when $\mathcal{B}_0$ is excitatory and $J^I = -gJ$ when $\mathcal{B}_0$ is inhibitory. Hence, the magnitude of ratio between the two SNRs for the two cases is almost $g = 7$ (the true ratio is actually smaller

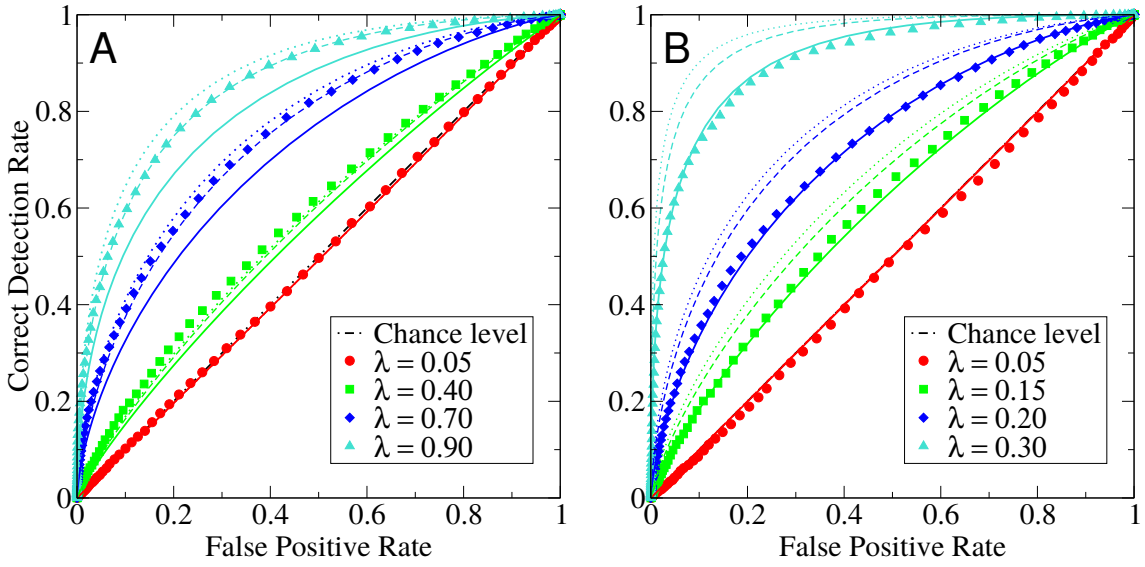

Figure 2.12. – **Receiver operating characteristic (ROC) curves for the detection of single-cell stimulation in the autonomous network.** **A**: excitatory $\mathcal{B}_0$. **B**: inhibitory $\mathcal{B}_0$. Simulations are indicated by symbols, theoretical estimates by lines: continuous lines indicate the "conservative" theory eq. (2.56), dashed lines the "average" theory eq. (2.54), and dotted lines the simplified "box-response" theory eq. (2.61). Parameters (**A**): $N_{\text{trials}} = 800$. Parameters (**B**): $N_{\text{trials}} = 600$. All other parameters are as in table 2.1. Each ROC curve is the average of sixteen realizations of the readout set. In evaluating eqs. (2.54), (2.56) and (2.61), the value of σ_A measured from numerical simulations was used, because the linear-response theory eq. (2.47) is not precise enough.

if nonlinearities and the effect of the recurrent connections are considered). It is worth noting that the stronger inhibitory weights were not chosen *ad hoc* to obtain a stronger effect for inhibitory cells, but as a necessary condition to achieve an asynchronous irregular spontaneous firing at low rates. Intuitively, because there are fewer inhibitory cells and their average firing rate is the same as that of excitatory cells (the two cell types are completely equivalent as far as their input and their dynamics are concerned), inhibitory spikes must be more powerful to keep the network firing rate low and stable.

Receiver operating characteristic curves

Although there are no experimental ROC curves to which simulation results can be compared, it is interesting to investigate them in the model, beginning with the case of the autonomous network, i.e. receiving no external shot noise. Figure 2.12 shows the ROC curves for four values of the detection bias λ, indicated in the inset. In this kind of plot, the diagonal corresponds to chance level (i.e. no detection), while the ideal detector would be represented by the upper edge of the graph (100% hits for any value of the false positive rate). Simulation results are

represented by symbols of various shapes and colors. The three types of lines represent the three theoretical approximations introduced in section 2.3.3: the conservative theory eq. (2.56) is plotted with continuous lines; the "average" theory eq. (2.54) is represented by dashed lines; and the simplified "box-shaped-response" theory eq. (2.62) is depicted with dotted lines.

The case of excitatory $\mathcal{B}_0$ is shown in fig. 2.12A. If neurons forming the readout population are chosen completely at random, ($\lambda = \lambda_0 = 0.05$) the ROC curve (red lines and circles) lies almost perfectly on the diagonal, which means that the perturbation is not detectable regardless of the choice of θ_d. However, for larger values of λ, i.e. if the readout population is biased towards $\mathcal{B}_1$, the ROC curve is well separated from the diagonal and the perturbation is detectable for any choice of the detection threshold that yields a false positive rate sufficiently far from zero or unity. The difference between the three theoretical approximations is quite modest for small λ, but increases for larger values of λ. As expected, the conservative theory (continuous lines) is always below the other two theories and it never exceeds the simulations. The simplified "box-response" theory (dotted lines) tends to overestimate the effect size. This theory is based on the approximation of the true shape of $\left\langle R^A_\lambda(t) \right\rangle$ with a box of equal area. However, eq. (2.48) depends non-linearly on the distance $\Delta R^A_\lambda(t) = \left\langle R^A_\lambda(t) \right\rangle - R_{\mathrm{sp}}$, so that jumping instantaneously to and from the maximal value produces a higher detection rate compared to true situation of a gradual transient. The "average" theory (dashed lines) is in-between the other two and agrees with simulations rather well.

Also in the case that $\mathcal{B}_0$ is inhibitory (fig. 2.12B), for unbiased readout ($\lambda = \lambda_0 = 0.05$) the ROC curve is hardly distinguishable from the diagonal. However, a small bias is sufficient to push the ROC away from the diagonal. Note that the values of λ shown here are different than in the previous case. For instance, green diamonds indicate here $\lambda = 0.15$ whereas they stand for $\lambda = 0.4$ in fig. 2.12A. Concerning the accuracy of the three theories in fig. 2.12B, there are some differences from the previous case (fig. 2.12A): here, simulations are in better agreement with the conservative theory (continuous lines), while all other theories overestimate the ROC curve; for large λ even the conservative theory slightly exceeds simulations.

Turning to the network receiving external shot-noise, the ROC curves resulting from the stimulation of an excitatory cell are displayed in fig. 2.13A, while the ROC curves relative to inhibitory $\mathcal{B}_0$ are shown in fig. 2.13B. The overall situation is quite similar to the previous case, although all ROC curves are here slightly closer to the diagonal if compared to the case that no external shot noise is present, i.e. the detectability is slightly reduced. The additional external noise reduces both the DC susceptibility and Δr_0, even if slightly (see fig. 1.9B on p. 28). The other terms appearing in eq. (2.67) remain substantially unchanged: the external noise has almost no effect on the cross-correlation term $S_{x_1x_2}$ (see also fig. A.9A) and causes a minor increase in the firing rate term s_{xx}. Altogether, these observations explain a moderate reduction in δ^A_λ and in the effect size.

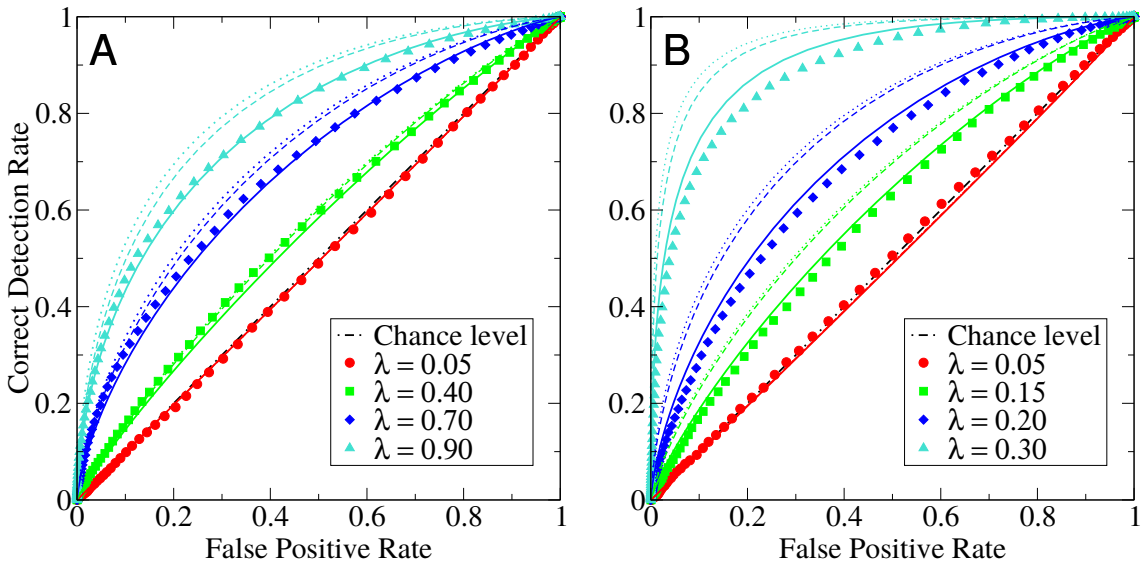

Figure 2.13. – **Receiver operating characteristic (ROC) curves for the detection of single-cell stimulation in the network with external shot-noise drive. A**: excitatory $\mathcal{B}_0$. **B**: inhibitory $\mathcal{B}_0$. Line and symbol coding as in fig. 2.12. Parameters (**A**): $N_{\text{trials}} = 900$. Parameters (**B**): $N_{\text{trials}} = 600$. All other parameters are as in table 2.2. Each ROC curve represents an average over sixteen realizations of the readout set.

One last observation is concerned with the position of the threshold. The effect size was defined in eq. (2.25) as the distance of the ROC curve from the diagonal for $\mathcal{FP} = 0.25$ (see also fig. 2.6). Although justified by the experimentally measured average value, the choice of this reference value for the false positive rate is still partially arbitrary. It is interesting to consider what the effect of changing the reference value is. By inspecting the ROC curves it is easy to see that reducing the reference value of $\mathcal{FP}$ would be beneficial only for ROC curves that are very well separated by the diagonal, i.e. for large values of λ and unrealistically large effect sizes. In the more realistic case of small λ, the maximal distance of the ROC from the diagonal is found at higher values of the false positive rates. Intriguingly, Houweling and Brecht (2008) observed that the effect size was larger for animals with higher responsiveness.

On the discrepancy of the theoretical approximation

In all previous plots, the disagreement between theory and simulations tends to be larger when $\mathcal{B}_0$ is inhibitory. The reasons behind this systematic difference are related to the fluctuations of the readout activity around its time-dependent mean, i.e.

$$R^A_\lambda(t) - \left\langle R^A_\lambda(t) \right\rangle , \tag{2.68}$$

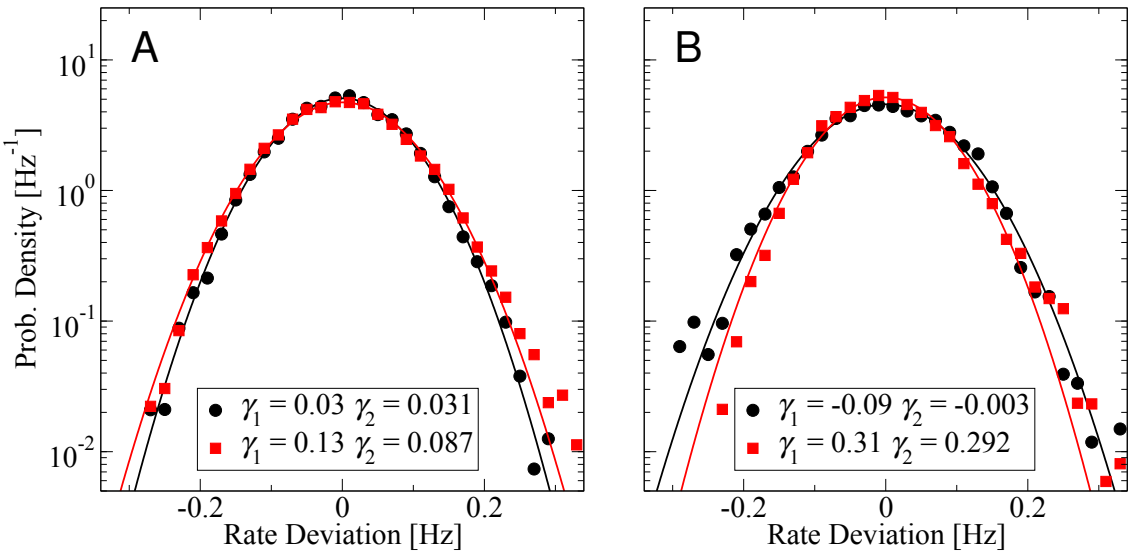

Figure 2.14. – **Fluctuations around the mean of the readout activity are less Gaussian during the stimulation, in particular when $\mathcal{B}_0$ is inhibitory (panel B).** Histograms of $R^A_\lambda(t) - \langle R^A_\lambda(t)\rangle$ with Gaussian fits. Black circles (histogram) and lines (Gaussian fit) refer to the spontaneous readout activity while the red squares and the red line represent the activity histogram and the Gaussian fit during stimulation, respectively. The legend reports the skewness γ_1 and the kurtosis γ_2. Parameters are as in table 2.1 and $\lambda = 1$. **A**: excitatory $\mathcal{B}_0$. **B**: inhibitory B_0. For the histogram of the activity during stimulation only the time window corresponding to the plateau of the mean value 200 ms $< t <$ 400 ms is considered.

where angular brackets indicate as usual average over trials. Figure 2.14 shows the histogram of the fluctuations around the trial-average eq. (2.68) together with a Gaussian fit in the spontaneous state and during the stimulation. The left graph fig. 2.14A refers to the case when $\mathcal{B}_0$ is excitatory and fig. 2.14B to the case of inhibitory $\mathcal{B}_0$. In the absence of the stimulus, the Gaussian distribution (black continuous line) represents an acceptable fit to the measured activity (black circles). Skewness (γ_1) and excess kurtosis (γ_2) of the data (reported in the plot legend next to the black dot) are also rather close to zero, confirming that the Gaussian approximation is reasonable. The theory assumes a Gaussian distribution both in the presence and absence of the stimulus and consider the variance as constant. Results in fig. 2.14 show that during the stimulation (red squares for the histogram and red lines for the Gaussian fit) the variance grows slightly if $\mathcal{B}_0$ is excitatory (fig. 2.14A) and shrinks a little if $\mathcal{B}_0$ is inhibitory (fig. 2.14B). What is more important, the stimulation affects the shape of the histogram in a different way, depending on which type of cell is stimulated: when $\mathcal{B}_0$ is excitatory, skewness and kurtosis grow only slightly; when $\mathcal{B}_0$ is inhibitory, the increase is considerably larger. It can be also clearly seen in fig. 2.14B that the left tail if the histogram becomes thinner and the right tail gains weight.

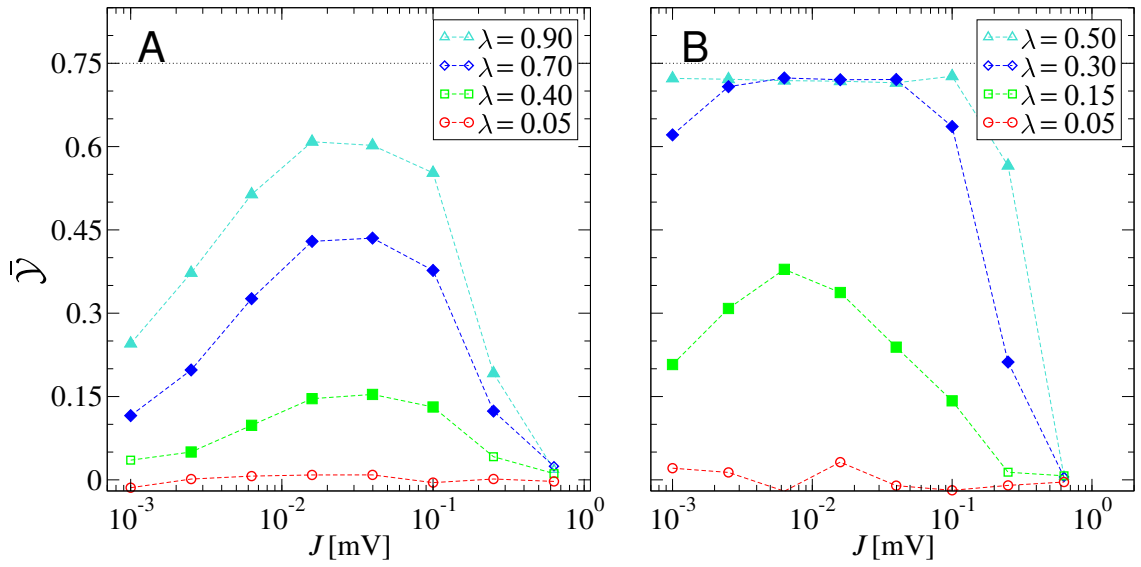

Figure 2.15. – The detectability of the single-cell stimulation shows a broad maximum for intermediate coupling. Effect size $\bar{\mathcal{Y}}$ as a function of the average coupling strength J for different overlaps λ. Closed symbols indicate data points significantly different than zero (p-value < 0.05). The black dotted line marks the maximal effect size $\bar{\mathcal{Y}} = 0.75$. **A**: Results for excitatory $\mathcal{B}_0$ obtained from $N_{\text{trials}} = 800$ trials. **B**: Results for inhibitory $\mathcal{B}_0$ obtained from $N_{\text{trials}} = 400$ trials. All parameters except J as in table 2.1.

The left tail is closer to the negative boundary and is, hence, more relevant for the detection of a negative deviation. Because the Gaussian fit overestimates the left tail, all theories predict a too large correct detection rate. The reason why the distribution of $R^A_\lambda(t)$ becomes more skewed to the right during the stimulation of an inhibitory $\mathcal{B}_0$ is simple rectification: firing rates cannot be negative, so that pushing the mean towards zero causes inevitably the histogram to lean towards positive values.

Optimal recurrent coupling strength

All results presented until this point were for an average excitatory coupling strength of 0.1 mV, which is a standard value in the literature because it is small enough for the diffusion approximation to work (it corresponds to one hundredth of the distance from reset to firing threshold), but it is not exceedingly small, if it is compared to experimentally measured values for postsynaptic potentials. The coupling strength can have a very strong influence on the spontaneous dynamics of the network and it is important to ascertain its effect on the detectability of the single-cell stimulation. To this end, the effect size will be plotted as a function of the average excitatory coupling J (the average inhibitory coupling is gJ), focusing first on the case of the autonomous network (no external white shot noise).

Figure 2.15A shows results for excitatory $\mathcal{B}_0$ and for the four values of the bias λ indicated in the inset. As in the previous cases, closed symbols indicate a statistically significant effect. If the readout is not biased ($\lambda = \lambda_0 = 0.05$, red circles) the effect is not significantly different from zero regardless of the coupling strength. For higher values of λ, the effect size displays a rather broad maximum in the intermediate coupling range. Also when the stimulated cell is inhibitory (Figure 2.15B) the single-cell stimulation is not detectable if the detector is not biased. However, even for a moderate bias, the range of couplings in which the stimulation is detectable spans more than two order of magnitudes.

Regardless of the type of the stimulated cell and of the value of λ, the effect size drops quite rapidly for $J > 0.1\,\mathrm{mV}$. As already discussed in section 2.2 when discussing the firing-rate response, increasing the coupling above a certain level leads to a qualitative change in the network noise (Wieland et al., 2015). In this dynamical regime, the spontaneous activity of the network has a lot of power at low frequencies, which, remembering eq. (2.65), make the denominator of the SNR grow.

For $J = 0$ the effect must be zero, because the perturbation cannot propagate to the rest of the network if neurons are not coupled at all (the stimulated cell is excluded from the readout set). Hence, $\bar{\mathcal{Y}}$ is expected to vanish when $J \to 0$. However, the decrease seen in fig. 2.15 is remarkably slow. The reason is that in an autonomous network the only noise source is generated by the recurrent connections and the noise intensity depends on the amplitude of the synaptic coupling. Recalling the approximation for the SNR

$$\delta^A_\lambda = \frac{\lambda\Delta r_1 + (1-\lambda)\Delta r_2}{\sigma_A} \approx \frac{\lambda\tau_m J^X \chi(0)\Delta r_0}{(s_{xx}/C + s_{x_1x_2})^{1/2}}, \tag{2.69}$$

the numerator is proportional to J. However, the consequences of the weakening noise affect the other parts of the SNR. Both the susceptibility $\chi(0)$ and Δr_0 grow (see also fig. 1.9A on p. 28), partially compensating the reduction in J. In the denominator, the term related to cross-correlations decreases. Although the spontaneous network firing rate increases because of the reduced recurrent inhibition (see fig. 2.3A), spike trains become very regular, as an effect of the weak noise. Because the low-frequency limit of the power spectrum is linked to the spike train regularity by the well-known relation eq. (1.17), the term s_{xx} also sinks. Altogether, these changes tend to compensate the effect of J so that the signal-to-noise ratio falls off very slowly.

The scenario in which the external shot-noise is present is considered in fig. 2.16, which shows the effect size $\bar{\mathcal{Y}}$ as a function of the *recurrent* coupling J: here, the intensity of the external noise J_{ext} is *not* varied along with J. Both when $\mathcal{B}_0$ is excitatory (fig. 2.16A) and when $\mathcal{B}_0$ is inhibitory (fig. 2.16B) the external noise generally reduces the effect size. However, while for strong and moderate coupling the change is minimal, a large difference can be seen in the low-coupling range: all curves fall off drop much faster for $J \to 0$ and the range in which the

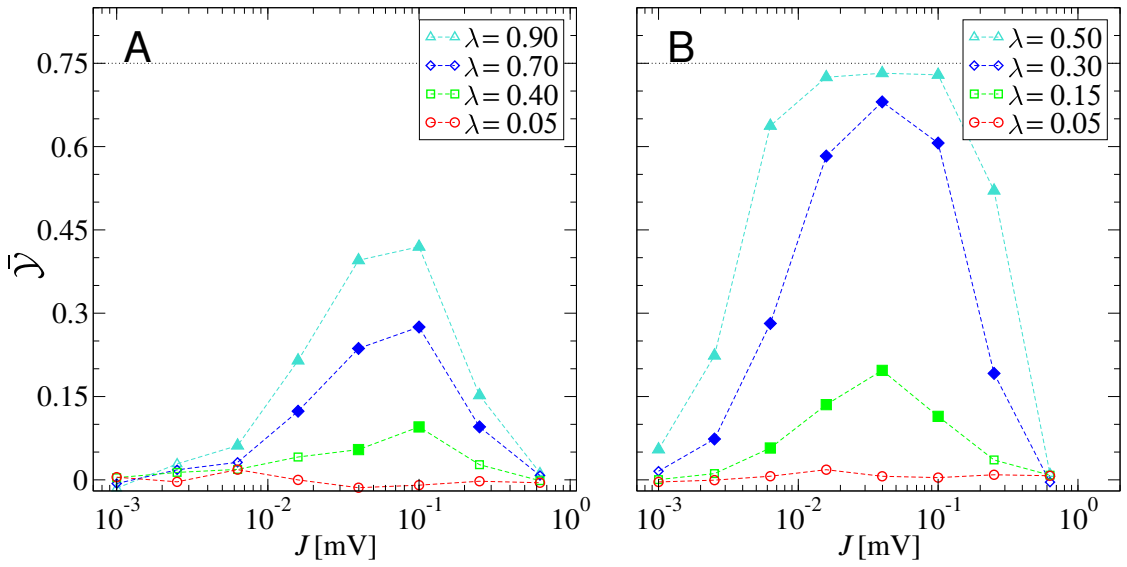

Figure 2.16. – The presence of external noise lowers the detectability mainly for weak recurrent coupling and narrows the peak at intermediate couplings. Effect size $\bar{\mathcal{Y}}$ as a function of the average coupling strength J for different overlaps λ for network with external input shot-noise. Closed symbols indicate data points significantly different than zero (p-value < 0.05). The black dotted line marks the maximal effect size $\bar{\mathcal{Y}} = 0.75$. **A**: Results for excitatory $\mathcal{B}_0$ obtained from $N_{\text{trials}} = 900$ trials. **B**: Results for inhibitory $\mathcal{B}_0$ obtained from $N_{\text{trials}} = 400$ trials. All other parameters except are as in table 2.2. Note that only the recurrent coupling J is varied, while the average amplitude of the external shot-noise J_{ext} is unchanged.

stimulation is detectable becomes narrower.

It has already been pointed out above that when $J \approx J_{\text{ext}}$, the main effect of the noise is a slight reduction of the DC susceptibility $\chi(0)$. For larger J, the effect of the external noise is essentially lost in the strong network noise. However, for weak recurrent coupling, the external noise has a twofold effect: first, it reduces the DC susceptibility $\chi(0)$; secondly, it restrains the denominator of the SNR from decreasing beyond a certain point. In particular, it prevents spike-trains to become perfectly regular and, hence, it avoids that $s_{xx} \to 0$.

2.5. Dependence on network size and robustness

In the following, the robustness of the core result of the last section will be tested, with particular emphasis on how the detectability depends on the network size. There are several possible ways of scaling the network. Therefore, to avoid the proliferation of possible cases and lose track of the main concepts, the analysis will be restricted to the detectability of an excitatory $\mathcal{B}_0$ in the presence of external shot noise, the case in which the detectability was most difficult.

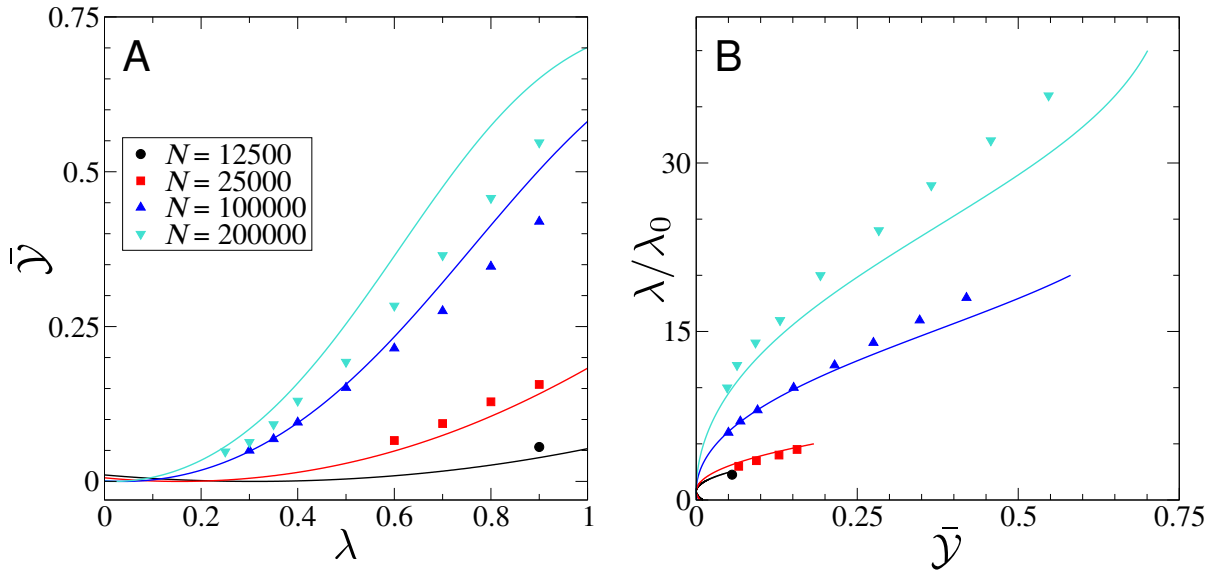

Figure 2.17. – **Scaling of the effect size for different network sizes and fixed number of connections per neuron.** **A**: Effect size as a function of the bias λ for different network sizes. Only significant data points (p-value$<$ 0.05) are shown. **B**: Minimal overlap relative to the natural value λ_0 necessary to obtain the effect size on the x-axis. The lines terminate at the maximal effect size achievable for each network size. Parameters as in table 2.2. Theoretical lines are computed from eq. (2.63), in which σ_A is measured from simulations.

Scaling with fixed number of connections per neuron

One possibility of scaling the network is to change its size N while keeping the number of connections per neuron C fixed. A consequence of this scaling is that the connection probability between any two selected neurons $p_c \approx C/N$ changes as $1/N$ when the network size N is varied. In particular, the network becomes sparser and sparser if the network size is increased.

Figure 2.17A shows the effect size as a function of λ for different network sizes when an excitatory neuron is stimulated. Non-significant data points are omitted to avoid overcrowding the left side of the plot. All theoretical curves display a minimum in the vicinity of $\lambda = \lambda_0 = C/N$ (the overlap corresponding to the unbiased readout) but increase very slowly for $\lambda < \lambda_0$. For $\lambda > \lambda_0$ all curves increase and reach their maximum for $\lambda = 1$. This maximum is an increasing function of the network size N, which can be explained by realizing that changing N has no strong influence on the firing rate or on the susceptibility of the neurons in the network. Hence, recalling eq. (2.69), the numerator of the SNR does not vary appreciably. However, the cross-correlations between neurons strongly depend on the connection probability $p_c = C/N$; in a larger, sparser network, cross-correlations are weaker, which makes the term $s_{x_1x_2}$ in eq. (2.69) decrease for increasing N and eventually leads to a larger SNR.

The results in fig. 2.17A may suggest the conclusion that increasing the size of the network is only beneficial to detection. However, constructing a readout with a given value of λ requires selecting neurons from $\mathcal{B}_1$ (whose size does not change) from an increasingly large network. In other words, the probability of finding by chance a neuron within $\mathcal{B}_1$ is smaller for larger N. One way of taking this fact into account is to consider the bias relative to the untrained case needed to achieve a given effect size. To this end, λ/λ_0 is plotted as a function of $\bar{\mathcal{Y}}$ (fig. 2.17B). All curves terminate at the maximal effect size achievable in the corresponding network size, which, as already known from fig. 2.17A, is larger for larger networks. However, provided that a given $\bar{\mathcal{Y}}$ can be obtained, the relative bias λ/λ_0 needed to achieve it is smaller for smaller networks.

In summary, this way of scaling the network imposes the following trade-off: enlarging the network raises the upper limit on the maximal effect size that can be attained; however, a smaller network requires less "learning" to change the bias to the value needed to obtain a given effect.

Scaling with fixed connection probability

Another way to scale the network is to keep the connection probability fixed. To this end, C and N must be proportional to each other so that varying the network size N is equivalent to varying the number of connections per neuron C and vice versa. Here, the network is inhibition-dominated so that varying the number of input connections while leaving other model parameters untouched would modify the strength of the recurrent inhibition; for this reason, enlarging the network would cause the spontaneous firing rate to decrease to zero. Typically, the average connection strength is adjusted to compensate for the changing number of inputs and avoid the network activity to become completely unlike the reference point. In the following, the mean recurrent coupling and the external shot-noise amplitude will scale as $\sim C^{-\alpha}$ and two cases will be considered: *strong synapses* ($\alpha = 1/2$) and *weak synapses* ($\alpha = 1$) (van Vreeswijk and Sompolinsky, 1998; Pehlevan and Sompolinsky, 2014). More precisely, let $\hat{J} = 0.1$ mV, $\hat{C} = 5000$, $\hat{C}_{ext} = 700$, and $\hat{g} = 7$ indicate the reference parameters; then, for a given number of inputs C, the network size is $N = C/p_c$, the average coupling strength and external shot noise amplitude are $J = J_{ext} = \hat{J} \cdot (\hat{C}/C)^{\alpha}$, and the number of external inputs is $C_{ext} = C \cdot (\hat{C}_{ext}/\hat{C})$.

We can now examine how the two scalings alter the mean input to each neuron. For weak synapses ($\alpha = 1$), the mean input to each neuron is

$$\mu_I = R_m I_{\text{ext}} + \tau_m \hat{J}\hat{C}\frac{1-\gamma g}{1+\gamma} r_{sp} + \tau_m r_{\text{ext}} \hat{J}\hat{C}_{\text{ext}},, \tag{2.70}$$

which does not depend on C and remains therefore constant independently of the network size. Conversely, in the case of strong synapses ($\alpha = 1/2$) the mean input reads

$$\mu_I = R_m I_{\text{ext}} + \tau_m \hat{J}(C \cdot \hat{C})^{\frac{1}{2}} \frac{1-\gamma g}{1+\gamma} r_{sp} + \tau_m r_{ext} \hat{J} \left(\frac{C}{\hat{C}}\right)^{\frac{1}{2}} \hat{C}_{ext}. \tag{2.71}$$

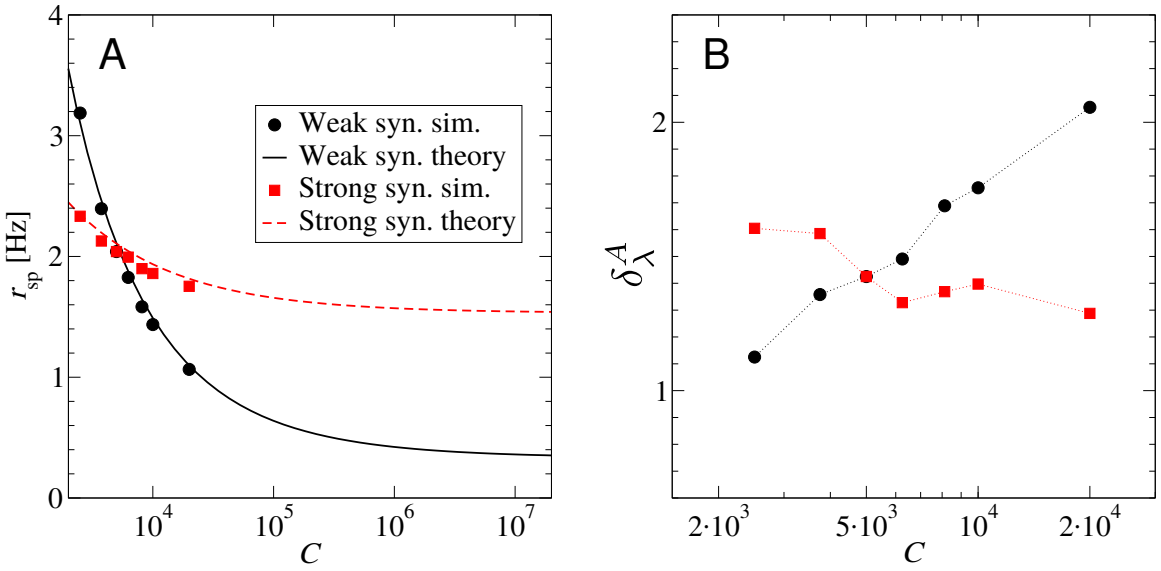

Figure 2.18. – **Qualitative behavior of the spontaneous network firing rate r_{sp} and of the signal-to-noise ratio δ^A_λ for increasing network size for the two different scalings of the synaptic coupling.** **A**: Spontaneous firing rate of the network for weak synapses (black circles, simulations; black continuous line, theory) and strong synapses (red squares, simulations; red dashed line, theory) as a function of the number of connections per neuron C (proportional to the network size $N = p_c C$). The spontaneous firing rate saturates at a non-zero for both scalings. **B**: Signal-to-noise ratio δ^A_λ as a function of the number of connections per neuron C. Non-scaled parameters are as table 2.3.

As seen in eq. (2.71), the mean of the external input as well as the prefactor multiplying the mean recurrent input, i.e. the strength of the recurrent feedback, would grow as $\sim C^{1/2}$. Setting

$$R_m I_{\text{ext}} = R_m \hat{I}_{\text{ext}} - \tau_m r_{\text{ext}} \hat{J} \left(\frac{C}{\hat{C}}\right)^{\frac{1}{2}} \hat{C}_{ext} \tag{2.72}$$

and

$$g = \frac{1}{\gamma}\left[1 + \left(\frac{\hat{C}}{C}\right)^{\frac{1}{2}} (\hat{g}\gamma - 1)\right] \tag{2.73}$$

compensates the change in the mean and recurrent feedback strength. Imposing that the mean input remains constant ensures that the spontaneous firing rate r_{sp} does not change too dramatically across different network sizes. Figure 2.18A shows the behavior of r_{sp} as a function of the number of inputs per neuron C (proportional to the network size N). For both scalings r_{sp} is a decreasing function of C. For strong synapses (red squares for simulations and dashed line for theory) the dependence is very weak (a variation of less than 20% over more than three orders of magnitude). For weak synapses (black circles and continuous line) the dependence is stronger. Importantly, the theory (which is in agreement with simulations for numerically

accessible network sizes) predicts saturation at a finite value for $C \to \infty$.

Having ensured that neither scaling produces a trivial limiting behavior for large networks, it is possible to inspect how the SNR changes in the range of network sizes that can be explored with simulations. For weak synapses (fig. 2.18B, black circles) the SNR increases moderately for growing network size, whereas for strong synapses it decreases slightly (fig. 2.18B, red squares). In fig. 2.18 a rather large overlap $\lambda = 0.79$ was chosen not to deal with very small values and ease the measurement of the SNR (simulations for the largest network size are computationally heavy, so that for this data point trials were limited to 200), but the trend is similar for other values of λ.

To gain some insight into the qualitative behavior of the SNR, its numerator and denominator can be inspected separately. According to eq. (2.66), the SNR reads

$$\delta^A = \frac{\Delta r_1}{\sigma_A} \approx \frac{\tau_m J \chi(0) \Delta r_0}{(s_{xx}/C + s_{x_1 x_2})^{1/2}}, \tag{2.74}$$

where $\lambda = 1$ was set for simplicity. Consider first the numerator Δr_1, plotted in fig. 2.19A. Because $J = \hat{J} \cdot (\hat{C}/C)^\alpha$, one could expect from eq. (2.74) a power-law decrease for Δr_1 with exponent $-\alpha$. Indeed, fig. 2.19A shows a power-law decrease for both scalings. However, the slope is not equal to $-\alpha$ in either case. For strong synapses (red squares), the numerator decreases as $\sim C^{-0.45}$ (and not $\sim C^{-0.5}$) and for weak synapses (black circles) the slope is $\sim C^{-0.55}$ (an even stronger difference from $J \sim C^{-1}$). In addition to simulations results, fig. 2.19 shows two theoretical lines for each case. The first theory (black continuous and red dashed line for weak and strong synapses, respectively) is based on the solution of eq. (2.14) and takes into account the recurrent feedback, as in all previous plots. The second theoretical line (dashed-dotted and dashed-double dotted line for weak and strong synapses, respectively) employs the further approximation $\Delta r_1 \approx \tau_m J \chi(0) \Delta r_0$. As both theories yield fairly similar results and agree with simulations, it be can concluded that the power-law can be understood in terms of the susceptibility $\chi(0)$ and of Δr_0, the change in the firing rate of B_0 induced by the signal. For strong synapses, both $\chi(0)$ and Δr_0 increase slightly, so that the exponent of the power-law decrease is somewhat larger than $-1/2$. For weak synapses, the susceptibility grows almost as fast as $J^{1/2}$ bringing the exponent of the power-law decrease again close to $1/2$. This increase in $\chi(0)$ is due to the system approaching a sort of rather pathological "threshold regime". In this regime, fluctuations in the input become smaller and smaller for larger C (and thus smaller J); to sustain fluctuation-driven firing, the minimum of the potential governing the dynamics of the membrane potential moves towards the firing threshold, which makes the system more sensitive to an external perturbation.

To study the behavior of the denominator of the SNR it is useful to resort to the decomposition

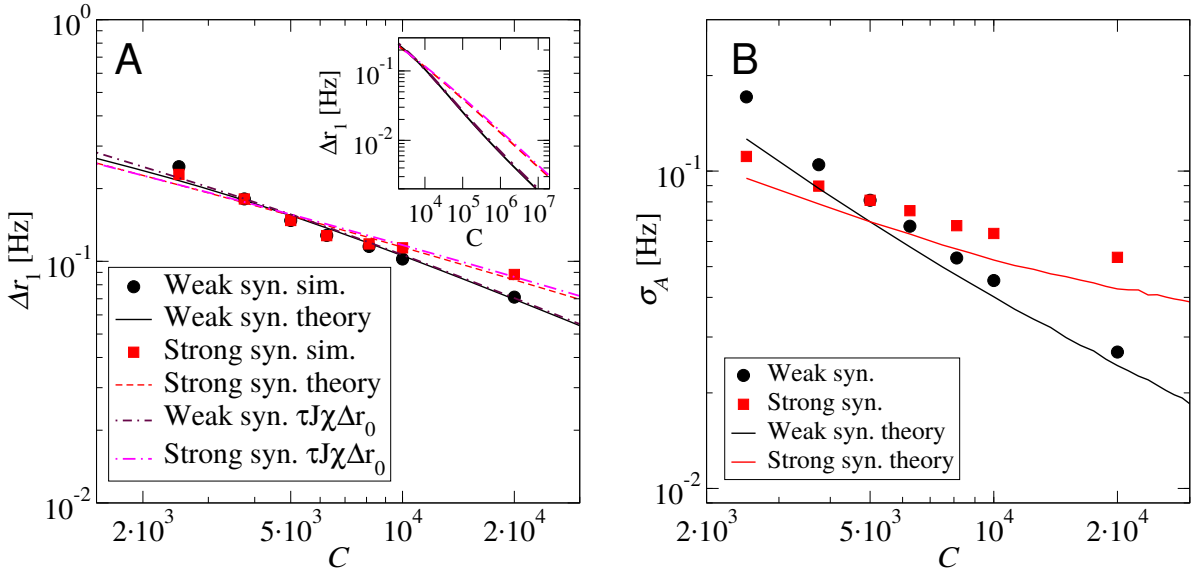

Figure 2.19. – **Qualitative behavior of numerator and denominator of the signal-to-noise ratio on increasing the network size for different scalings of the synaptic coupling.** **A**: Numerator of the SNR. Simulation, full self-consistent solution as in eq. (2.14) and approximation in eq. (2.74) as indicated in the legend (inset shows theory over a larger range). The agreement of dashed and solid lines illustrates that recurrent loops have only a minor effect on the numerator of the SNR. **B**: Standard deviation of the readout activity σ_A as a function of C for the two considered weight scalings. For large values of C, numerical instabilities in the computation of the double integrals in eq. (2.11), the shot-noise theory for the power spectrum S_{xx}, flaw the evaluation of the theory. To avoid this problem, the approximation $S_{xx} \approx r_{\mathrm{sp}}$ was made in the theory shown above. For very large values of C, even the numerical integration of the single integral in eq. (A.6), the DC susceptibility $\chi(0)$, becomes unstable causing the small fluctuations seen in the theory at the right end of the plot. Constant parameters and reference values are as in table 2.2, see main text for definition of parameters scaled together with the network size (which is here proportional to C).

of σ_A^2 resulting from eq. (2.65), reported here for convenience:

$$\sigma_A^2 \approx \frac{S_{xx}(0)}{C\sqrt{\pi}\tau_f} + \frac{S_{x_1x_2}(0)}{\sqrt{\pi}\tau_f} = \frac{s_{xx}}{C} + s_{x_1x_2}. \tag{2.75}$$

As discussed in section 2.3.2, the shot-noise theory provides a good approximation for the single spike-train power spectrum S_{xx}, whereas the linear-response theory

$$S_{x_1x_2}(f) \approx S_{xx}p_c\left(|A|^2\frac{C_E(1+g^2\gamma)}{|1-AC_E(1-g\gamma)|^2} + 2\Re\left[\frac{A}{1-AC_E(1-g\gamma)}\right]\frac{1-g\gamma}{1+\gamma}\right) \tag{2.76}$$

underestimates the low-frequency limit of the cross-spectrum $S_{x_1x_2}(0)$. Still, using the shot-noise theory for S_{xx} together with eqs. (2.75) and (2.76) succeeds in predicting the qualitative behavior of σ_A for both weak and strong synapses, as shown in fig. 2.19B.

For weak synapses, the first term in eq. (2.75) decays as $\sim C^{-1.5}$, because the firing rate (proportional to the power-spectrum) decays as $\sim C^{-0.5}$ (see fig. 2.18A). The second term falls off with a power law with a similar exponent, because $A = \tau_m J\chi\mathcal{D} \sim C^{-0.55}$, as discussed above for the numerator. Therefore, as long as the susceptibility keeps growing with C, the denominator decreases faster than the numerator, i.e. with an exponent $\sim C^{-0.75}$ versus $\sim C^{-0.55}$, which explains the increase in the SNR seen in fig. 2.18B.

For strong synapses, the firing rate is rather constant (see fig. 2.18A), so that the first term in eq. (2.75) simply decays as $\sim C^{-1}$. The susceptibility increases only weakly in this scaling, so that most dependencies of the second term in eq. (2.75) are weak; only the term $(1+g^2\gamma)$ decreases significantly at first but settles on the value $1+1/\gamma$ as C becomes large. Therefore, after an initial decrease, σ_A should saturate. In fig. 2.19B it can be seen that σ_A falls off slower for strong synapses and the slope diminishes for increasing C. Although the saturation is not visible for network sizes that can be simulated, even in the range shown in fig. 2.19B the denominator decreases slower than the numerator, so that the SNR becomes smaller for increasing network size.

Asymptotic "thermodynamic" limit

Although the network sizes considered above are already quite large, it is interesting to investigate the asymptotic behavior when $N \to \infty$, i.e. in the "thermodynamic" limit. In the first scenario considered above, the number of connections was left fixed while varying the network size. In this case, the thermodynamic limit corresponds to an infinitely sparse network, that is, $p_c \approx C/N \to 0$. A first consequence of $p_c \to 0$ is that cross-correlations between neurons vanish. Hence, from eq. (2.75) it follows that $\sigma_A^2 \to s_{xx}/C$. As far as the numerator of the SNR is concerned, in the limit $N \to \infty$ the stimulation does not affect the firing-rate of $\mathcal{B}_2$, which stays at its unperturbed value r_{sp}, as letting $p_c \to 0$ in eqs. (2.14) and (2.15) shows. The reason

is that the effect of $\mathcal{B}_1$ on the firing rate of $\mathcal{B}_2$ depends on the probability of connections from $\mathcal{B}_1$ to $\mathcal{B}_2$, which is C/N and vanishes in the limit $N \to \infty$. In the end, the signal-to-noise ratio for a fixed λ settles at

$$\delta^A(\lambda) \xrightarrow{N\to\infty} \frac{\Delta r_1(\lambda)}{(s_{xx}/C)^{1/2}} \approx \frac{\lambda \tau_m J \chi(0) \Delta r_0}{(s_{xx}/C)^{1/2}}, \tag{2.77}$$

which would imply a finite effect size for any given $\lambda > 0$. However, the natural overlap $\lambda_0 = C/N$ vanishes for $N \to \infty$ as well, which causes the relative bias necessary to achieve any non-zero effect size to diverge.

In the second scenario considered above, in which p_c is constant and C is proportional to N, two cases were distinguished. In the case that synapses are weak ($J \sim C^{-1}$), the noise intensity goes to zero for $C \to \infty$, which means that, at some point, the noise-driven firing cannot be sustained anymore. When neurons enter the mean-driven regime and the susceptibility saturates (the f-I curve becomes that of a deterministic LIF neuron), so does the term related to the spike train power-spectrum, i.e. $s_{xx} \to s_{xx,\infty}$. This argument suffices to infer the limiting behavior of the SNR:

$$\delta^A \sim \frac{C^{-1}}{(s_{xx,\infty}C^{-1} + s_{x_1x_2})^{1/2}} < \frac{C^{-1}}{(s_{xx,\infty}C^{-1})^{1/2}} \sim C^{-\frac{1}{2}} \xrightarrow{C\to\infty} 0. \tag{2.78}$$

The theory predicts that cross-correlations vanish as $1/C$ for $C \to \infty$. However, the term $s_{x_1x_2}$ can simply be neglected as long as it is positive because it would only drive the SNR faster towards zero. For all network sizes, the dominant contribution to the cross-spectrum term is given by the first term in eq. (2.76), which is positive, so that it is reasonable to assume that $s_{x_1x_2}$ remains positive while it decreases to zero.

In the other case of strong synapses ($J \sim C^{-1/2}$), the theory predicts that the low-frequency limit of the cross-spectrum saturates at a finite value:

$$\frac{S_{x_1x_2}(0)}{\sqrt{\pi}\tau_f} \xrightarrow{C\to\infty} \frac{S_{xx,\infty}(0)}{\sqrt{\pi}\tau_f} p_c |A_\infty|^2 \hat{C} \frac{1+\gamma^{-1}}{1+\gamma} = s_{x_1x_2,\infty}. \tag{2.79}$$

As observed above, for the standard parameters the saturation of $S_{x_1x_2}$ is very slow. In a smaller network with a smaller value for the reference value $\hat{C} = 1000$ the plateau is reached faster, as seen in fig. 2.20. With a non-vanishing cross-correlation term, nothing can prevent the SNR to drop to zero as $C^{-1/2}$

$$\delta^A \sim \frac{C^{-0.5}}{(s_{x_1x_2,\infty})^{1/2}} \sim C^{-0.5} \xrightarrow{C\to\infty} 0. \tag{2.80}$$

In conclusion, detecting the single-cell stimulation in a network of infinite size turns out to be problematic in all considered scenarios. Taking the thermodynamic limit without scaling the number of connections per neuron (the infinitely sparse network) is often done (Gerstner et al., 2014), because when $p_c \to 0$ cross-correlations between neuron vanish, which makes the mean

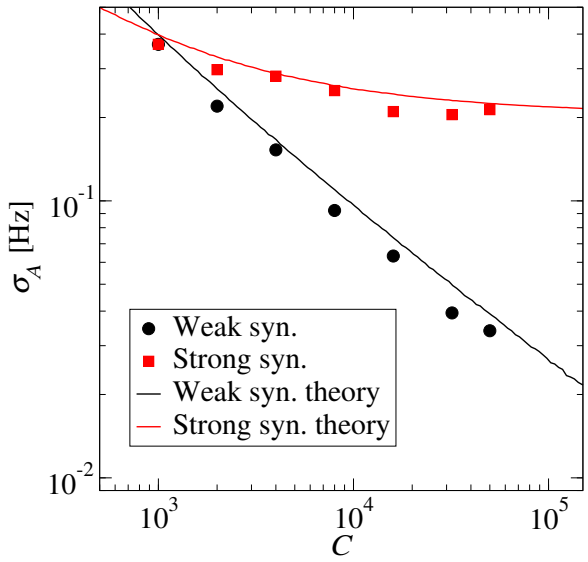

Figure 2.20. – Faster saturation of the readout fluctuations as a function of C for smaller reference value for the number of connections. Standard deviation of the readout activity as a function of C for a smaller, denser network. Parameters are $\hat{C} = 1000$, $p_c = 0.2$, $C_{\text{ext}} = 1000$, $R_m I_0 = -2\,\text{mV}$, otherwise as in table 2.2.

field description via a Fokker-Planck equation easier. In this case, assuming that λ can be set to a value larger than zero, the single-cell stimulation would be detectable in the thermodynamic limit and the effect would be larger than for any finite size of the network. However, one may wonder whether it is possible to train the readout to find a set of finite size in an infinitely large network. If the ratio λ/λ_0 is interpreted as a measure for the amount of learning necessary to achieve the readout bias, the divergence caused by $\lambda_0 \to 0$ may be interpreted as the fact that an "infinite training" would be required to detect the stimulus. When the connections probability is kept constant and the number of connections per neuron C is let vary proportionally to the network size, it is natural to assume that the readout set can also be scaled in the same fashion, which solves the problem of constructing a biased readout set. In this scenario, to prevent the network from being completely silent in the limit $C, N \to \infty$, the synaptic connections need to be reduced as C increases, as discussed above. Two cases were considered. Scaling the synaptic strength as $\sim 1/\sqrt{C}$ (the case of strong synapses) is typical of "balanced network models" (van Vreeswijk and Sompolinsky, 1996, 1998). The idea of this scaling is that the variance of the input fluctuations should stay constant regardless of the network size, if cross-correlations are neglected. Ironically, it is the non-vanishing cross-correlation term $s_{x_1 x_2,\infty}$ that is responsible for the decay of the SNR to zero. Scaling the synaptic strength as $\sim 1/C$ (the case of weak synapses) was considered by Pehlevan and Sompolinsky (2014) to model a network with vanishing input fluctuations in the thermodynamic limit. In this case, however, the numerator of the SNR decays

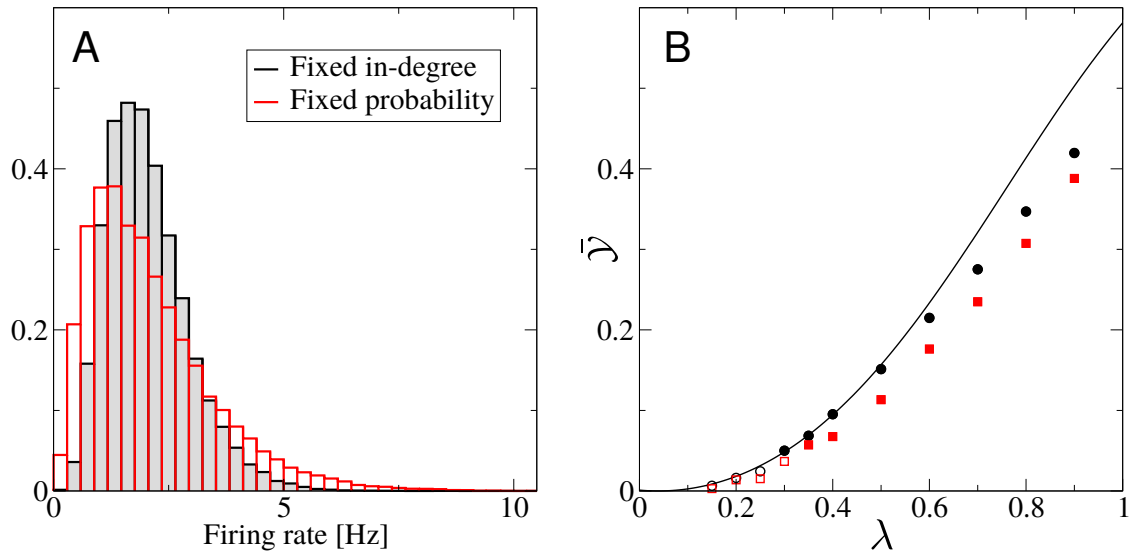

Figure 2.21. – **Comparison between fixed and distributed (Erdős-Rényi topology) in-degree.** **A**: Firing rate histograms for the two topologies. **B**: Effect size for the two topologies. Black circles and continuous line refer to the fixed in-degree. Red squares are simulation results for the (Erdős-Rényi topology). Closed symbols indicate data points significantly different from zero (p-value < 0.05). Parameters are as in table 2.2.

faster than the uncorrelated fluctuations, so that δ^A is again expected to vanish.

These considerations suggest that the ability to detect the single-cell perturbation is to be regarded, formally, as a finite-size effect, which is a theoretically interesting finding. However, it can not be a counterargument to the general validity of the model, as cortical networks are not infinitely large and the perturbation is detectable for a wide range network sizes.

Detection for a network with Erdős-Rényi topology

One simplifying assumption made in the definition of the network model is the fixed in-degree, i.e. that each neuron receives exactly the same number of input connections, which is obviously not the case in real cortical networks. Here, the impact of a random in-degree on the detectability of the single-cells stimulation is studied. The simplest way of randomizing the in-degree is to consider all possible (directed) neuron pairs independently of each other and to create a link between them with a fixed probability. The resulting connectivity matrix is known as an Erdős-Rényi graph, in which in-degrees are not fixed but binomially distributed. One first consequence of a random in-degree on the network dynamics is shown in fig. 2.21A: the spread of the firing rate distribution is wider for the Erdős-Rényi network (red histogram) compared to the network with fixed in-degree (gray histogram). It is not surprising that increasing the heterogeneity in the network input to each neuron leads to an increment in the heterogeneity of the firing rate.

The effect of this increased heterogeneity in the network on the detectability is to lower the effect size, as shown in fig. 2.21B (red squares for Erdős-Rényi graph, black circles for fixed in-degree). However, the difference is modest and the perturbation remains detectable.

Optimal readout filter time

Until now, the readout filter time constant was set to $\tau_f = 100\,\text{ms}$ and never changed. Here, the effect of changing τ_f will be inspected. The numerator of the SNR, as defined in eq. (2.66), does not explicitly depend on τ_f, while eq. (2.65) states that the variance of the readout activity is inversely proportional to τ_f,

$$\sigma_A^2 = \frac{1}{\tau_f\sqrt{\pi}}\left(\frac{S_{xx}(0)}{C} + S_{x_1x_2}(0)\right), \tag{2.81}$$

which may suggest that expanding the readout time scale τ_f can only improve the SNR and, thus, the effect size. The simulation results in Figure 2.22A show that the effect size (black circles) is indeed an increasing function of τ_f, at first. However, above 300 ms it rapidly drops.

To understand why the effect size is no simple increasing function of the time filter, one must reexamine to the assumptions under which it is reasonable to consider the numerator of the SNR as independent of τ_f. The numerator of the SNR is related to the time-dependent mean of the readout activity. As explained in section 2.3.2, after the stimulus onset at $t = 0$, the time-dependent mean of $R_\lambda^A(t)$ deviates from its spontaneous value according to

$$\left\langle \Delta R_\lambda^A(t) \right\rangle = \Delta r_\lambda \Delta a(t) * \mathcal{F}_{\tau_f}(t), \tag{2.82}$$

where the time course $\Delta a(t) * \mathcal{F}_{\tau_f}(t)$ is given by eq. (2.28) and is plotted in fig. 2.22B for three values of τ_f (the y-axis is rescaled so that $\Delta r_\lambda = 1$). For the standard choice of $\tau_f = 100\,\text{ms}$ (and for smaller values), the maximum value is reached and maintained for a time that is similar to duration of the stimulus. In such a situation, it is acceptable to approximate the time course of $\langle \Delta R_\lambda^A(t) \rangle$ with a box-shaped function of length T_w and height Δr_λ. For larger values of τ_f, however, this approximation is not justified, because there is no plateau and the peak value of $\langle \Delta R_\lambda^A(t) \rangle$ is well below Δr_λ. One possibility to account for the time-dependence of the readout activity in the theory would be to use one of the two theoretical approximations eq. (2.53) or eq. (2.56), thus waiving the simple relationship between the effect size and SNR eq. (2.63) shown in fig. 2.10. Alternatively, the definition of the SNR can be changed to reflect the effect of a long filter time constant on the mean readout activity:

$$\delta_\lambda^A = \frac{\Delta \hat{r}_\lambda}{\sigma_A}, \tag{2.83}$$

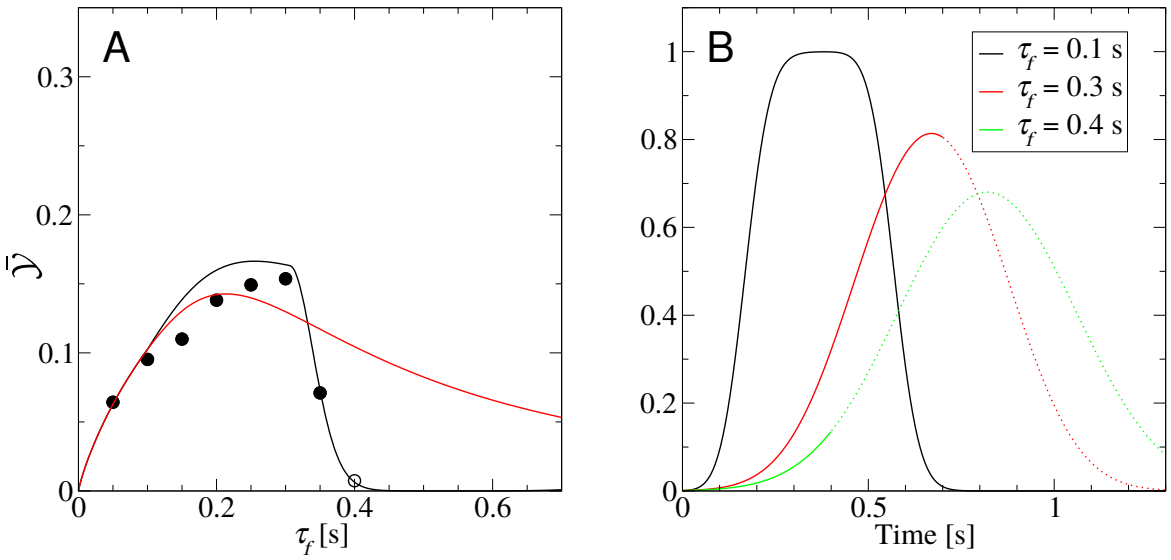

Figure 2.22. – **Optimal readout time filter constant for detection A**: Effect size as a function of readout filter time constant τ_f. Circles indicate simulations and closed symbols mark values significantly different than zero ($p < 0.05$). Black continuous lines denote the theoretical prediction based on the combination of eq. (2.83) with eq. (2.63). For both theory and simulations the detection time window T_w is reduced so that the constraint eq. (2.84), i.e. $T_w < T/2 - \tau_f$ is fulfilled. If, instead, T_w is left fixed and the simulation time T is changed to fulfill eq. (2.84), the red theoretical line is obtained. For this case, simulations are not available. All other parameters as in table 2.2. **B**: Time dependence of the mean readout activity, i.e. $\Delta a(t) * \mathcal{F}_{\tau_f}(t)$, for three values of τ_f, computed from the theoretical expression in eq. (2.28). The x-axis ends at the standard value of T_w, so that the entire range shown is used in evaluating the red theoretical line in panel A. For simulations and the corresponding black theoretical line, the dotted portion of the curve falls outside T_w and is discarded.

where $\Delta\hat{r}_\lambda$ indicates the maximum value of eq. (2.82) within the time window for detection, i.e. for $0 < t < T_w$. If this definition of the SNR is used together with the theory eq. (2.63), the red continuous line shown in fig. 2.22A is obtained. Although the theoretical approximation in red correctly captures the non-monotonic dependence of $\bar{\mathcal{Y}}$ on τ_f, the quantitative agreement is not satisfactory because the theory predicts a maximum around 200 ms and a gradual decline, whereas simulations peak around 300 ms and then fall quite abruptly to zero. This discrepancy has a technical explanation related to the limited simulation time and is explained as follows. The support of readout filter is an interval of length of $3\tau_f$. Hence, if the total simulation time is T, the total duration of the time series obtained for the readout activity is $T - 3\tau_f$. The reason is that the causality of the filter requires using the first $3\tau_f$ ms to compute the first time point, which implies that only $T - 3\tau_f$ of the total simulation time is available for the detection experiment. Because T is centered around $t = 0$, where the stimulus is turned on, the total

simulation time in the spontaneous state is $T/2-3\tau_f$. Two detection windows of *equal length* are required for an unbiased measurement of false positive and hit rates. To this end, the following constraint must hold

$$T_w < \frac{T}{2} - 3\tau_f. \tag{2.84}$$

In other words, using a longer filter time requires either an increase in the total simulation time or a shorter detection time window. The network considered here is large and simulating a large number of trials for longer time windows is impractical. Therefore, instead of increasing T, T_w was reduced such that eq. (2.84) is satisfied. Figure 2.22B visualizes the portion of signal cut by the shortened detection window as dashed line. For $\tau_f = 100\,\text{ms}$ (black line) the downswing of the signal reaches zero well before the end of the detection time window (the plot range ends is the standard value of T_w). For $\tau_f = 300\,\text{ms}$ (red line), the detection window cuts off the signal just before its maximum value, which explains why the maximal detectability is in the vicinity of this value: the portion of signal most decisive for the detection (around the peak value) is kept, while the part after the downswing (which is basically indistinguishable from the catch trial situation) is cut off. For $\tau_f = 400\,\text{ms}$, the signal is cut before the peak value is reached, which explains the drop in the effect size. Taking the constraint eq. (2.84) into account when calculating the SNR in eq. (2.83) leads to the theoretical prediction plotted as black continuous line in fig. 2.22, which (except for the usual slight overestimation for non-small effect sizes) agrees well with simulations.

To summarize these observations, increasing the filter time constant τ_f has the beneficial effect of averaging spontaneous fluctuations over a larger time window, thus reducing the readout noise and favoring detectability. However, it has also the detrimental effect of smearing the response to the signal over a larger time interval, which eventually exceeds the detection time window thus hindering detectability. The combination of these two effects leads to the existence of an optimal time constant.

2.6. Detectability for the single-barrel network and balanced input

All results presented up to this point concerned a rather large network, corresponding to a patch of somatosensory cortex of about 0.5 mm (Meyer et al., 2010), which would extend over about five barrels. The choice not to limit the network size to a single barrel can be supported by the fact that the anatomical separation between barrels is only pronounced for recurrent connections within layer 4 and for input connections from the thalamus (see section 1.1), while the stimulated cells were neither chosen exclusively from layer 4 nor from the center of barrels, but from various depths and potentially from the *septum*, the area between barrels. On the other hand, barrels do form a functional unit and it could be assumed that some degree of separation seeps to other layers because of the preferred vertical orientation of cortical axons. For this reason, it is interesting to consider a network of the size of one single barrel as alternative setting for the single-cell detection virtual experiment, which is done in this final part of the chapter.

The average size of one barrel is roughly $N = 20000$ neurons. Maintaining a connection probability pf $p_c = 0.05$ implies that each neuron receives now input from $C = 1000$ randomly selected neurons (of which $C_E = 800$ are excitatory and $C_I = 200$ are inhibitory) within the barrel. Here, the external shot-noise input does not model only the sensory input from the thalamus or top-down projections from other distant cortical areas, but also input from the surrounding cortical cells. Hence, it consists of both excitatory and inhibitory input spikes, with the usual ratio of four to one. More precisely, in this section the external input reads:

$$I_{\mathrm{ext},k}(t) = I_0 + \frac{\tau_m}{R_m}\left[\sum_{j=1}^{C_{\mathrm{ext}}}\sum_{l} J_{k,j,l}\delta(t - t_{k,j,l}) - g\sum_{p=1}^{\gamma C_{\mathrm{ext}}}\sum_{q} J_{k,p,q}\delta(t - t_{k,p,q})\right]. \tag{2.85}$$

In the last equation, the spike times are realizations of independent Poisson processes and the weights are drawn independently from an exponential distribution with mean J, as before. The rate of the γC_{ext} inhibitory Poisson processes is set for simplicity equal to the excitatory one $r_{\mathrm{ext}} = 10\,\mathrm{Hz}$. The number of external inputs is chosen such that $C_E + C_{\mathrm{ext}} = 4000$. The presence of inhibitory shot-noise permits to achieve a low spontaneous firing rate with smaller inhibitory weights. Two scenarios are considered: in the first one, the recurrent input is slightly inhibition-dominated with relative inhibitory strength $g = 4.5$; in the second one, the recurrent input is "balanced", i.e. $g = 1/\gamma = 4$. The constant mean input was adjusted to obtain a spontaneous firing rate similar to the previous sections, i.e. $r_{\mathrm{sp}} \approx 2\,\mathrm{Hz}$, and is different in the two cases. All parameters used in this section are summarized in table 2.3.

As mentioned in section 2.3.2 and discussed in more detail in appendix A, with this choice of parameters the linear-response theory for network cross-correlations is in fairly good agreement with simulations in the low-frequency range (see fig. A.9 on p. 223). Hence, the theoretical lines shown here do not require measuring σ_A from simulations and stem from a fully analytical

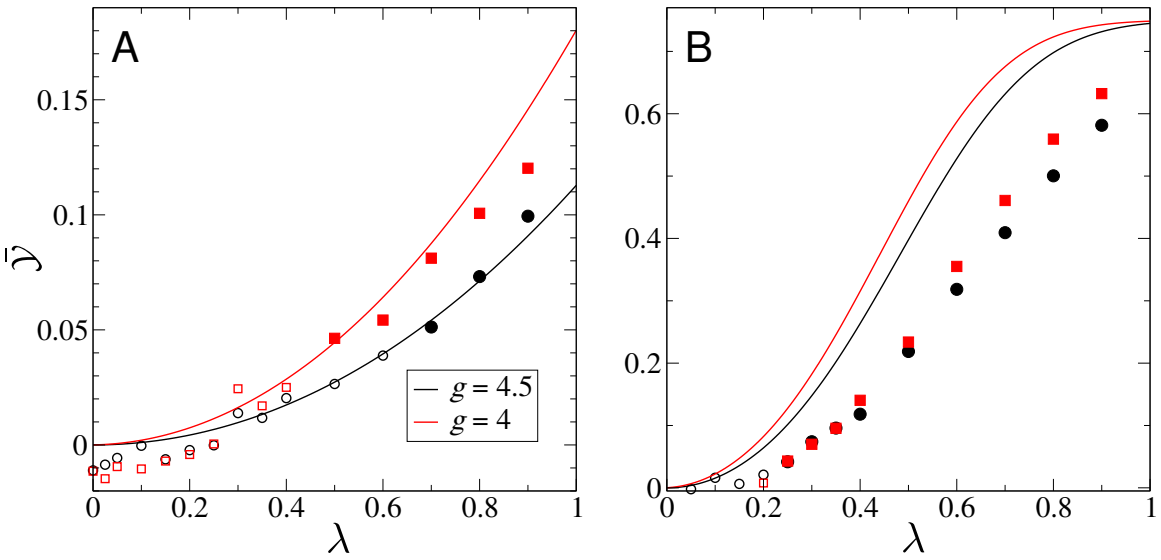

Figure 2.23. – **Detectability of single-cell stimulation for the "single-barrel" network.** Effect size $\bar{\mathcal{Y}}$ as a function of the bias λ for excitatory (**A**) and inhibitory (**B**) $\mathcal{B}_0$. The case of prevalently inhibitory recurrent input $g = 4.5$ is represented by black circles for simulations and black lines for the theory. The case of balanced recurrent input $g = 4$ is displayed by red squares for simulations and red lines for theory. Other parameters as in table 2.3.

calculation.

The effect size as a function of the detection bias λ is shown in fig. 2.23. In the case that $\mathcal{B}_0$ is excitatory (fig. 2.23A), the effect size is globally smaller than for the larger network considered in the previous sections. The maximum effect size is here about $\bar{\mathcal{Y}} \approx 0.1$ for the inhibition-dominated case $g = 4.5$ (black circles for simulations, black lines for theory) and slightly higher for the balanced case $g = 4$ (red squares for simulations and red lines for theory). The smallest bias that results in a statistically significant effect size (represented, as usual, by closed symbols) is $\lambda = 0.7$ for $g = 4.5$ and $\lambda = 0.5$ for $g = 4$. When the stimulated cell is inhibitory (fig. 2.23B), the effect size is larger. The difference between the two cases $g = 4.5$ and $g = 4$ is appreciable only for rather large values of the bias and in both cases the minimum bias required for a statistically significant detection is $\lambda = 0.25$.

The smaller effect sizes observed here - if compared to larger network of the previous sections - is due to a decrease in the numerator and to an increase in denominator of the SNR. The former is caused by the larger amount of external noise that lowers the single-neuron susceptibility. The latter can be understood by first recalling the decomposition of the readout noise variance introduced in eq. (2.65)

$$\sigma_A^2 = \frac{1}{\tau_f\sqrt{\pi}}\left(\frac{S_{xx}(0)}{C} + S_{x_1x_2}(0)\right) = \frac{s_{xx}}{C} + s_{x_1x_2}, \tag{2.86}$$

and then analyzing separately the magnitude of the two terms. The low-frequency limit of the average cross spectrum $S_{x_1x_2}(0)$ is slightly larger here (see fig. A.9B) than in the case of larger network (see fig. A.9A). The term s_{xx} is proportional to the single-neuron power spectrum and, thus, similar in the two cases. However, the size of the readout set, C, is here four times as small, so that the first term is four times larger here than in the previous sections.

As to the difference between the two values of the inhibition strength g, it is usually argued that perfect balance between excitation and inhibition is beneficial for signal transmission (Vogels and Abbott, 2009) and decorrelation of spiking activity (Renart et al., 2010). Here, the slightly larger detectability observed for $g = 4$ is attributable to a larger susceptibility that overcompensates the increase seen in the cross-correlation term (compare the two cases in fig. A.9B).

2.7. Summary and discussion

The aim of this chapter was to test the "network modeler's null hypothesis" on the problem posed by the experiment by Houweling and Brecht (2008): is there a way of detecting the stimulation of a single-cell in a large random network, whose spontaneous activity is asynchronous, irregular, and at low firing rates, as in the barrel cortex? To answer this question, the Amit-Brunel network was chosen as the network model. After showing that the shot-noise theory by Richardson and Swarbrick (2010) can be used self-consistently to approximate the firing rates both in the spontaneous state (section 2.1) and in response to the single cell stimulation (section 2.2), a comparatively simple stimulus detector was introduced in section 2.3. This detector reacts to deviations from the spontaneous state in the activity of a readout subpopulation that can be biased towards $\mathcal{B}_1$, the set of neurons receiving direct input from the stimulated cell $\mathcal{B}_0$. In the same section, a theory was developed to understand what features of the network's dynamics affect the detectability of the stimulation and to estimate the effect size analytically. The theory highlighted the role of single-neuron susceptibility and of cross-correlations as the crucial factors influencing the detectability and it helped to interpret how the effect size depends on the parameters of the model.

The main result of the chapter is that the stimulation is detectable if the readout is biased

(section 2.4). In accordance with the experiments, an inhibitory cell is more easily detectable than an excitatory one. If the bias is interpreted as a caricature for the training that the animals undergo before the detection task, the fact that the effect size is not significant for no bias, i.e. that the system must be somehow prepared, is again consistent with the experiment, because naive subjects cannot report the occurrence of the single-cell stimulation. Because the defining property of a network is, as a matter of fact, the interaction among its elements, the strength of this interaction is perhaps the most representative parameter of a network. The results in section 2.4 show that the detectability of the stimulation quickly drops when the recurrent coupling strength increases beyond the "critical" value that marks the point where the network noise undergoes a qualitative change (Ostojic, 2014; Wieland et al., 2015). When the coupling is weak, the external noise plays an important role: in the absence of it, the stimulation is detectable for extremely small values of the coupling strength, whereas its presence prevents the detectability. In an intermediate range, the presence of the external noise has a very moderate influence on the effect size and the stimulation detectability reaches an optimal value.

Cortical networks consist of a large but finite number of neurons. Still, one may speculate whether the stimulation is detectable in the "thermodynamic" limit of an infinitely large network. Although the answer was found to be negative for all considered ways of scaling the network, the results in section 2.5 show that the stimulation is detectable for any network size in a biologically plausible range. In the same section, the robustness of the results was further tested by relaxing the assumption of fixed in-degree and by varying the time scale of the readout. Intuitively, the detectability reaches an optimum when the readout filter constant is long enough to average fast fluctuations but not so long that the signal is also averaged out.

If the network model is shrunk to the size of a single cortical barrel, the single-cell is still detectable, although the necessary bias was found to be somewhat larger (section 2.6). The two principal causes are the larger noise in the readout activity, due to the smaller size of the readout population, and the reduced susceptibility of the single neurons caused by the strong external noise.

As mentioned above, the bias of the detector, λ, was interpreted here as a representation for the effect of the training phase. In particular, it can represent a change in the synaptic connections taking place not within the recurrent network itself, but in the connections from the barrel cortex to the readout area. Another possibility would be to model learning by changing the recurrent connections within the network, for instance by creating hub cells. One potential objection to this approach, however, is that the identity of specific cells should not play a role during learning, because the training phase is based on extracellular microstimulation (see section 1.2) that does not target single cells but a whole area. As a matter of fact, in the purely random network considered here, the overlap of the readout set with $\mathcal{B}_1$ (which is equal to the bias λ) depends on the identity of $\mathcal{B}_0$ as well. Therefore, changing $\mathcal{B}_0$ without changing the

readout set is tantamount to a change in the overlap λ, which has two important consequences. On the one hand, averaging over different values of λ is a way to represent the average over different cells. On the other hand, if λ is changed at random, the probability of the new λ being large enough to permit detection is rather small in almost all of the considered scenarios. How could the training phase affect the readout bias, so that many cells with a sufficient λ can be found at random? One possibility is that the microstimulation marks a preferred area for the readout, thus increasing the average λ for all cells in the vicinity. This hypothesis is consistent with the fact that the single-cell stimulation is only detectable in the same area where the training was carried out, although a true notion of vicinity is only possible in a network with a spatial structure.

The detector was equipped with two thresholds symmetrically placed around the mean spontaneous readout activity R_{sp}. Although this definition permits the detection of both an excitatory and an inhibitory $\mathcal{B}_0$, it also presents drawbacks. From the technical point of view, it requires a precise knowledge of R_{sp} to function properly: if the detector is misaligned, one of the two barriers becomes less effective (a detailed explanation of this problem is found in appendix B). It is not realistic to assume that a perfect estimate of R_{sp} must always be available to the detector, so that this weakness is not merely technical, but also conceptional. Furthermore, the biological meaning of the two barriers is problematic: crossing an upper boundary can be viewed as the activity of an excitatory population reaching a level high enough to trigger a downstream effect, while crossing a lower boundary can be seen as the activity of an inhibitory population being transiently so low that its target is disinhibited. However, a mixed population playing both roles at the same time is more difficult to interpret. Finally, in this model the "decision" about the presence of the stimulus is taken within the stimulated network itself, which is not biologically plausible in a primary sensory area.

2.8. Tables of parameters

All parameters used in this chapter are listed here in three tables for reference. Table 2.1 displays parameters for the "standard autonomous network" (sections 2.1 to 2.4). The other two tables quote only parameters that are different from those in table 2.1. Table 2.2 refers to the case of the "standard driven network" (sections 2.1 to 2.5). Finally, parameters values used in section 2.6 (the "single barrel network") are shown in table 2.3.

Symbol	Value	Description
τ_m	20 ms	membrane time constant
τ_{ref}	2 ms	refractory period
v_T	20 mV	threshold voltage
v_R	10 mV	reset voltage
$R_m I_0$	22 mV	constant external input
C_{ext}	0	number of excitatory Poisson inputs per neuron
$C_{\text{ext,inh}}$	0	number of inhibitory Poisson inputs per neuron
r_{ext}	0 Hz	rate of external Poisson inputs
N_E	80 000	number of excitatory neurons in the BCN
γ	0.25	ratio of inhibitory to excitatory neurons
N_I	γN_E	number of inhibitory neurons
C_E	4000	number of excitatory inputs per neuron
C_I	γC_E	number of inhibitory inputs per neuron
J	0.1 mV	average synaptic coupling strength (exponentially distributed)
g	7	relative strength of inhibitory to excitatory coupling
D_{min}	0.5 ms	minimum of uniform transmission delay distribution
D_{max}	2.0 ms	maximum of uniform transmission delay distribution
T_s	400 ms	stimulus duration
$R_m \Delta I_{\text{ext}}$	23 mV	stimulus intensity
T_w	1300 ms	time window for single-cell detection
τ_f	100 ms	width of time filter for detection
N_A	$C = (1+\gamma)C_E$	number of neurons in the readout set $\mathcal{S}^A$
T_{ic}	400 ms to 1200 ms	initial simulation time to forget initial conditions
T	3000 ms	simulation time (data acquisition)
Δt_{sim}	0.1 ms	simulation time step
N_{trials}	600 to 900	Number of trials

Table 2.1. – Numerical values of all parameters for the "standard autonomous network". The initial simulation time T_{ic}, used to forget initial conditions and discarded from the data acquisition, was adapted to the strength of the recurrent coupling, where the maximal value is chosen for the minimal coupling and the minimal value for couplings around the "critical" value $J \approx 0.2$ mV because, as shown by Wieland et al. (2015), the autocorrelation time of the network activity has a minimum in the critical range.

Symbol	Value	Description
$R_m I_0$	5.2 mV	constant external input
C_{ext}	700	number of excitatory Poisson inputs per neuron
r_{ext}	12 Hz	rate of external Poisson inputs

Table 2.2. – Numerical values of all parameters for the "standard driven network". Only parameters that differ from table 2.1 are listed.

Symbol	Value	Description
$R_m I_0$	14 mV or 8.2 mV	constant external input ($R_m I_{\text{ext}} = 8.2\,\text{mV}$ when $g = 4$)
C_{ext}	3200	number of excitatory Poisson inputs per neuron
$C_{\text{ext,inh}}$	800	number of inhibitory Poisson inputs per neuron
r_{ext}	10 Hz	rate of external Poisson inputs
N_E	16 000	number of excitatory neurons in the BCN
C_E	800	number of excitatory inputs per neuron
g	4.5 or 4	strength of inhibition relative to excitation
N_A	$C = (1+\gamma)C_E$	number of neurons in the readout set $\mathcal{S}^A$
N_{trials}	1000	Number of trials

Table 2.3. – Numerical values of all parameters for the "single barrel network". Only parameters that differ from table 2.1 are listed.

Chapter 3.

Detection of Single-Cell Stimulation by a Second Readout Network

The previous chapter was successful in proving that the "modeler's null hypothesis" on the experiment by Houweling and Brecht (2008) can be - to some extent - verified by using a comparatively simple detection procedure. Despite this achievement, the model of chapter 2 suffers from several drawbacks, discussed in section 2.7. Some of them are rather technical in nature, others are more general. One rather important limitation is that the detector decision on the presence of the stimulus is taken by analyzing the activity of a subset of the stimulated network itself. Because the barrel cortex belongs to the primary sensory cortex, it is perhaps unlikely that the activation of a population within it represents a conscious decision able to trigger a complex motor reaction.

In this chapter, the previous model will be revisited and extended in several ways. The most important addition will be a second network receiving input from the stimulated one and acting as a readout. The performance of this new readout model as a detector will be compared to the previous one for different scenarios. Central results are that the simplest configuration for the readout network - a population of integrator neurons with no recurrent connections - yields an effect size roughly equivalent to that of the previous chapter, whereas a readout network with recurrent local inhibition is much more effective in detecting the single-cell stimulation.

The chapter is structured as follows. Section 3.1 introduces the problem and describes the general structure of the new model. The three following sections examine three possible configurations for the new readout network and show how effective each of them is in detecting the single-cell stimulation: section 3.2 considers a single readout population of neurons without recurrent connections; in section 3.3, the readout network is provided with local recurrent and feed-forward inhibition but no recurrent excitation; in section 3.4, the readout is performed by a fully recurrent excitatory-inhibitory (E-I) network. The final section 3.5 summarizes and discusses all the results of the chapter in more detail.

3.1. General model

The way the single-cell stimulation experiment by Houweling and Brecht (2008) is mimicked here is essentially the same as in chapter 2. The focus of this chapter is the comparison of two readout schemes to detect the single-cell stimulation, which are portrayed schematically in fig. 3.1. Both readout schemes receive input from the same network model representing the portion of the barrel cortex surrounding the stimulated cell. This network model is once more a large (10^5) Amit-Brunel network tuned to fire in the asynchronous irregular regime and at low rates. Although it does not possess features specific to a particular cortical area, for brevity it will be referred to as the "barrel cortex network" (BCN).

In the readout scheme A (fig. 3.1A), a set of neurons $\mathcal{S}^A$ is selected at random but with a bias towards the set of neurons (labeled $\mathcal{B}_1$) receiving a direct connection from the stimulated cell. The filtered activity of $\mathcal{S}^A$ is read by a detector that reacts whenever the activity hits a decision threshold. The performance of the detector can be characterized by systematically varying the position of the threshold (the detector's sensitivity) to obtain the correct detection rate as a function of the false positive rate, i.e. the receiver operating characteristic (ROC) curve. The final output of the readout scheme is the *effect size* $\bar{\mathcal{Y}}$, defined as the difference between the correct detection rate and the false positive rate for a suitably chosen sensitivity of the detector. Besides some minor differences discussed in the following, this readout scheme is the one used in chapter 2.

In the readout scheme B (fig. 3.1B), a second network of leaky integrate-and-fire (LIF) neurons, "the readout network" (RN), receives feed-forward input from the BCN. These feed-forward connections from the BCN to the RN can also be biased towards $\mathcal{B}_1$. The summed activity of all excitatory neurons in the RN is low-pass filtered and fed to the detector, which works in the same way as in the readout scheme A. As in the previous chapter, the readout bias is indicated by $0 \leq \lambda \leq 1$, where $\lambda = \lambda_0 = C/N$ is the unbiased case, $\lambda > \lambda_0$ indicates a bias towards $\mathcal{B}_1$, and $\lambda < \lambda_0$ corresponds to a bias against $\mathcal{B}_1$. In the rest of this section, the components of the model are described in more detail. All numerical parameters used in the chapter are listed in table 3.1 on p. 132.

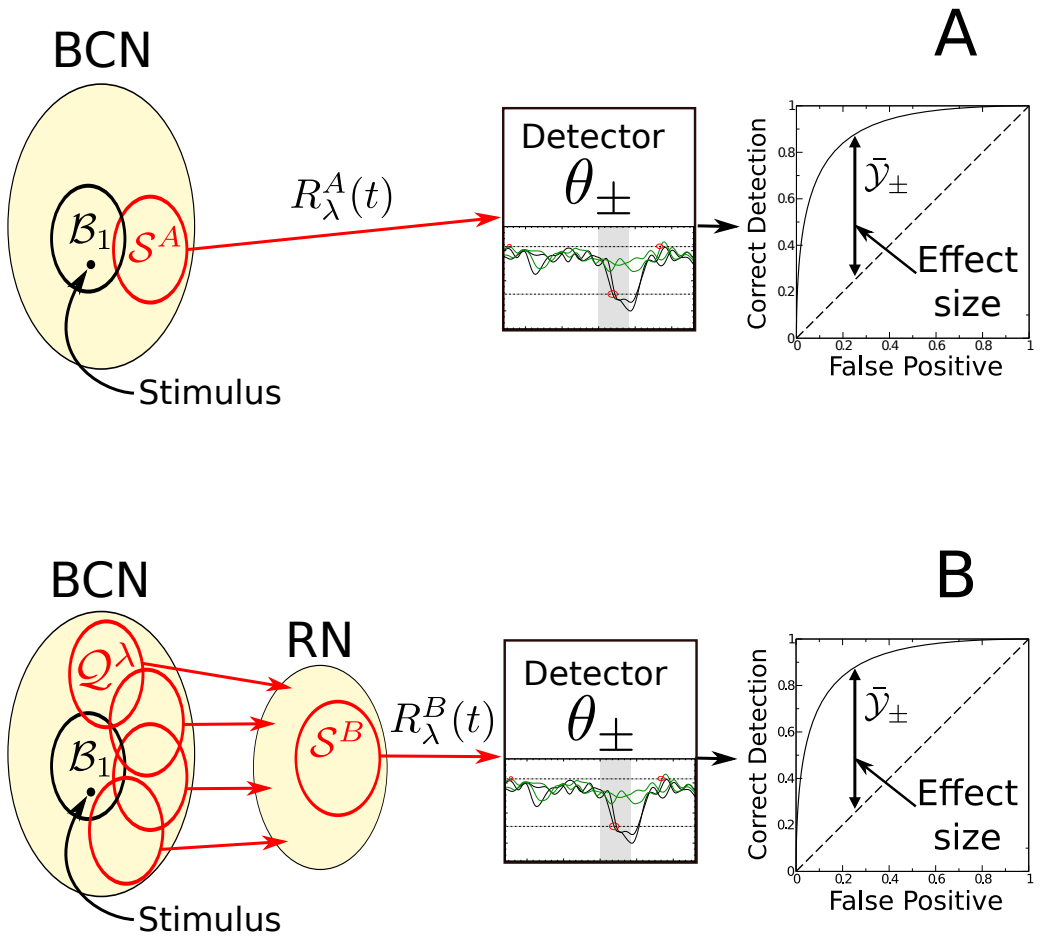

Figure 3.1. – **Illustration of the two detection schemes compared in this chapter**. A cell selected at random in the "Barrel Cortex Network" (BCN) is stimulated. The detection scheme A is analogous to the one considered in chapter 3, except for a minor difference in the detector, that is now provided with a single barrier instead of two. In the detection scheme B, a second readout network (RN) receives feed-forward input from the BCN. The activity of all excitatory neurons within the readout network, $\mathcal{S}^B$, is the input to the detector. In both cases, the parameter λ quantifies the bias of the readout towards the set of neurons $\mathcal{B}_1$, the neurons of one synapse away from the stimulated neuron. Three different architectures for the RN are considered.

3.1.1. Barrel cortex network and single-cell stimulation

The BCN is identical to the "standard driven network" considered in the previous chapter. Hence, this part of the model is only briefly reviewed here for convenience; the reader is referred to section 2.1 for a detailed description of the model and for a discussion of the network's spontaneous dynamics.

The BCN is composed of $N_E = 80\,000$ excitatory and $N_I = \gamma N_E = 20\,000$ LIF neurons. All neurons evolve according to

$$\tau_m \frac{\mathrm{d}v_k}{\mathrm{d}t} = -v_k + R_m[I_{\mathrm{ext}}(t) + I_{\mathrm{syn,k}}(t)] \tag{3.1}$$

with the usual fire-and-reset rule (see section 1.4). Neurons are coupled by current-based synapses (see section 1.4), so that the recurrent input term is

$$I_{\mathrm{rec},k}(t) = \frac{\tau_m}{R_m}\left[\sum_{j\in\mathcal{P}_e(k)} J_{kj}x_j(t-D_{kj}) - g\sum_{\ell\in\mathcal{P}_i(k)} J_{k\ell}x_\ell(t-D_{k\ell})\right], \tag{3.2}$$

where $\mathcal{P}_e(k)$ is a set of $C_E = 4000$ randomly selected excitatory neurons, $\mathcal{P}_i(k)$ is a set of $C_I = \gamma C_E = 1000$ randomly chosen inhibitory neurons, J_{kj} and $J_{k\ell}$ are independent exponentially distributed random couplings with mean $J = 0.1\,\mathrm{mV}$, D_{kj} and D_{kl} are transmission delays drawn from a uniform distribution in the interval 0.5 ms to 2.0 ms. Autapses (self-coupling) are excluded, i.e. $J_{ii} = 0$ for any i. As in the previous chapter, the connection probability between two neurons is approximately independent of the neuron type and is sparse $p_c \approx C/N = 0.05$. The external input is the sum of a constant term I_0 and of Poissonian shot noise

$$I_{\mathrm{ext},k}(t) = I_0 + \frac{\tau_m}{R_m}\left[\sum_{j=1}^{C_{\mathrm{ext}}}\sum_l J_{k,j,l}\delta(t-t_{k,j,l})\right], \tag{3.3}$$

where $t_{k,j,l}$ are independent spiking times with mean rate $r_{\mathrm{ext}} = 12\,\mathrm{Hz}$, $C_{\mathrm{ext}} = 700$ is the number of external inputs per neuron, and $J_{k,j,l}$ are independent samples from an exponential distribution with mean $J = 0.1\,\mathrm{mV}$. The constant input term $R_m I_0 = 5.2\,\mathrm{mV}$ is chosen such that the total mean external input is slightly above the threshold voltage, which keeps the network firing. However, because the mean recurrent input is inhibitory, the *total* mean input is below the threshold and neurons fire driven by input fluctuations. As a consequence, the spontaneous state of the network is asynchronous and irregular; the mean firing rate is low $r_{\mathrm{sp}} \approx 2\,\mathrm{Hz}$.

The single-cell stimulation experiment is modeled as in the previous chapter. First, to forget the random initial conditions, the network is run for $T_{\mathrm{ic}} = 500\,\mathrm{ms}$, which are discarded from the analysis. Then, it is simulated for a further time interval of $T = 3\,\mathrm{s}$, centered on $t = 0$. A cell is selected at random as the site for the stimulation. As in the previous chapter, the

juxtacellular current injection is mimicked by raising the constant input term by $\Delta R_m I_0 = 23\,\mathrm{mV}$ for $0 < t < T_s = 400\,\mathrm{ms}$.

In the experiment by Houweling and Brecht (2008), each cell was stimulated about 15 times on average, but the effect size was averaged over many (51) cells from different animals. Results of this chapter are based on a similar total number of trials (900), but for simplicity the realization of the network is changed in each trial. Therefore, a trial-average is here equivalent to an average over the network topology (including weights and delays), over the external input noise, and over random initial conditions, unlike the previous chapter, in which the network topology was kept fixed.

As in the previous chapter, three subsets of quite unequal size will be distinguished within the BCN (see fig. 2.3): the first subset is $\mathcal{B}_0$, which is simply the stimulated cell; the second one is $\mathcal{B}_1$, the set of neurons receiving direct input from $\mathcal{B}_0$, which has average size $N_1 = Np_c$; the third one is $\mathcal{B}_2$, which consists of all other neurons and counts therefore $N_2 = N - N_1 - 1$ neurons, on average. The firing-rate response of the three populations was discussed in section 2.2.

3.1.2. Readout

In both readout schemes, the detector receives input from the set of neurons $\mathcal{S}^X$, where $X = A, B$ indicates one of the two schemes. In the readout scheme A, ($X = A$, illustrated in fig. 3.1A), the readout set $\mathcal{S}^A$ is a subset of the BCN constructed by randomly picking $\lambda\hat{C}$ excitatory neurons from $\mathcal{B}_1$ and $(1-\lambda)\hat{C}$ excitatory neurons from $\mathcal{B}_2$ ($\mathcal{B}_0$ is excluded from $\mathcal{S}^A$). The size of $\mathcal{S}^A$ is $N_A = \hat{C} = 4000$ except for section 3.4, in which $N_A = \hat{C} = 1000$. The readout set is constructed almost in same way as in the previous chapter, but here only excitatory neurons are used to construct $\mathcal{S}^A$. By this construction, λ is the overlap between $\mathcal{S}^A$ and $\mathcal{B}_1$.

In the readout scheme B, ($X = B$, depicted in fig. 3.1B), the readout set $\mathcal{S}^B$ consists of all excitatory neurons within the RN. The size of $\mathcal{S}^B$ is $N_B = 10\,000$. Each neuron in the RN receives feedforward input from the BCN, as well as shot-noise, and input from within the RN. The connections from the BCN to the RN are only excitatory, because long-range axons usually originate from excitatory neurons and axons of inhibitory neurons are typically confined to a local area (Helmstaedter et al., 2009a,b; Tremblay et al., 2016). More precisely, each neuron in the RN evolves according to

$$\tau_m \frac{\mathrm{d}v_k}{\mathrm{d}t} = -v_k(t) + R_m[I_{\mathrm{ext},k}(t) + I_{\mathrm{rec},k}(t) + I_{\mathrm{FF},k}(t)] \tag{3.4}$$

with the fire-and-reset rule. The first input term $I_{\mathrm{ext},k}(t)$ is analogous to eq. (3.3) and represents the input from other areas. The term $I_{\mathrm{rec},k}(t)$ is the input from within the RN. The third term

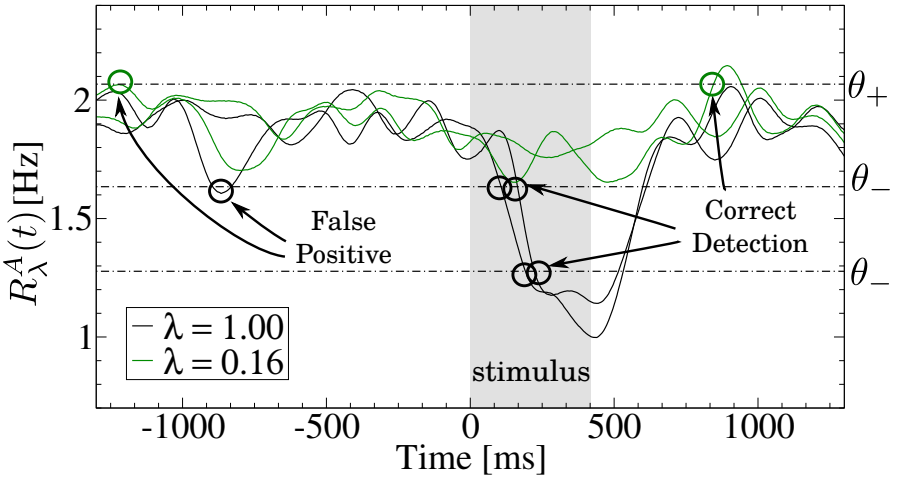

Figure 3.2. – **Working principle of the single-barrier detector used in this chapter.** Two example realizations of the readout activity $R^A_\lambda(t)$ are plotted as continuous lines for two values of the bias λ. A detection event is registered when the activity exceeds the upper-barrier detector θ_+ or falls below the lower-barrier detector θ_-. If the crossing occurs for $t \in (-T_w, 0)$ a *false positive* event is registered. If the crossing takes place in the interval $(0, T_w)$, it counts as *correct detection* (a hit).

$I_{\mathrm{FF},k}(t)$ models the input from the BCN to the RN and reads

$$I_{\mathrm{FF},k}(t) = \frac{\tau_m}{R_m} \sum_{j \in \mathcal{Q}_\lambda(k)} J_{kj}^{\mathrm{FF}} x_j(t - D_{kj}^{\mathrm{FF}}), \tag{3.5}$$

where the weights J_{kj}^{FF} and delays D_{kj}^{FF} are randomly distributed in the same way as for the BCN. The set of neurons $\mathcal{Q}_\lambda(k)$ comprises $\hat{C}$ excitatory neurons selected at random from $\mathcal{B}_1$ with probability λ, otherwise taken from $\mathcal{B}_2$ (consistently with the readout scheme A, $\mathcal{B}_0$ is left out). Put differently, λ is here the *average* overlap between $\mathcal{B}_1$ and $\mathcal{Q}_\lambda(k)$, the neurons presynaptically connected to the readout network. The overlap for each single $\mathcal{Q}_\lambda(k)$ is binomially distributed with relative standard deviation $\sigma_\lambda/\lambda = \sqrt{(1-\lambda)/\hat{C}}$, which is quite small because the number of feed-forward inputs per neuron $\hat{C}$ is large. The expression for the term $I_{\mathrm{rec},k}(t)$ depends on the particular architecture of the RN and is therefore discussed separately in the respective section.

3.1.3. Detector and effect size

The detector is essentially the same in the two readout schemes and is rather similar to the detector of chapter 2.

First, the readout activity $R_\lambda^X(t)$ is obtained by filtering the average activity of the readout population $\mathcal{S}^X$:

$$R_\lambda^X(t) = \frac{1}{N_X} \sum_{j \in \mathcal{S}^X} x_j(t) \star F_{\tau_f}(t), \tag{3.6}$$

where N_X is the number of neurons in $\mathcal{S}^X$ and $\star$ indicates convolution. The filter is the same as in chapter 2, a truncated Gaussian

$$F_{\tau_f}(t) = \frac{H(t)H(3\tau_f - t)}{\sqrt{\pi \tau_f^2/2}} \exp\left[\frac{-(t - 3\tau_f/2)^2}{\tau_f^2/2}\right] \tag{3.7}$$

of width $\tau_f = 100\,\mathrm{ms}$, shifted to ensure causality. The difference to chapter 2 is that the single detector with two symmetric barriers is replaced here by two detectors with a single barrier. One detector responds to crossings of an upper detection boundary θ_+, the other one to crossings of a lower barrier θ_- (fig. 3.2). As in the previous chapter, there are two detection time windows: the first one is $(-T_w, 0)$, in which a threshold crossing defines a false positive; the second one is $(0, T_w)$, in which the detector's response is considered a correct detection. Therefore, whenever the readout activity exceeds θ_+ for $t \in (-T_w, 0)$, the θ_+ detector records a false positive event (for convenience, the same symbol will be used to indicate both the detector and the corresponding threshold). Averaging over trials yields the false positive rate for the θ_+ detector:

$$\mathcal{FP}_\lambda^+(\theta_+) = \left\langle \max_{t \in (-T_w, 0)} \left\{ H(R_\lambda^X(t) - \theta_+) \right\} \right\rangle, \tag{3.8}$$

where $H(t)$ is the Heaviside function. Analogously, whenever the readout activity falls below the level θ_- in the time window $(-T_w, 0)$ the θ_- detector marks a false positive. The false positive rate for the θ_- detector is then

$$\mathcal{FP}^-_\lambda(\theta_-) = \left\langle \max_{t\in(-T_w,0)} \left\{ H(\theta_- - R^X_\lambda(t)) \right\} \right\rangle. \tag{3.9}$$

Correct detection rates $\mathcal{CD}^\pm_\lambda(\theta_\pm)$ are defined in the same way but for $t \in (0, T_w)$. The effect size is defined as the difference between correct detection rate and false positive rate for the threshold $\bar{\theta}_\pm$ that corresponds to a false positive rate of 25%

$$\bar{\mathcal{Y}}^\pm(\lambda) = \mathcal{Y}^\pm_\lambda(\bar{\theta}_\pm) = \mathcal{CD}^\pm_\lambda(\bar{\theta}_\pm) - \mathcal{FP}^\pm_\lambda(\bar{\theta}_\pm), \tag{3.10}$$

with the threshold $\bar{\theta}_\pm$ obeying

$$\mathcal{FP}^\pm_\lambda(\bar{\theta}_\pm) = 0.25. \tag{3.11}$$

In the readout scheme A, detection rates are averaged over sixteen realizations of $\mathcal{S}^A$, as in the previous chapter. In the readout scheme B there is no need for such average because each of the $N_B = 10\,000$ neurons in $\mathcal{S}^B$ receives input from a different subset of neurons in the BCN so that the possible subsets of the BCN are already adequately sampled. The statistical significance of the effect size is computed by using Fisher's exact test in the same way as in chapter 2. More details on the procedure and on subtle issues due to the averaging on realizations of $\mathcal{S}^A$ can be found in section 2.3.

3.1.4. Detection theory and signal-to-noise ratio

The theory to estimate the effect size discussed in section 2.3.3 can be applied to this detector with minor modifications. Hence, the formulas used to plot the theoretical lines of the next sections are derived below without an in-depth explanation. As in the previous chapter, a central role in the discussion is played by the *signal-to-noise ratio* (SNR) for the two readout populations ($X = A, B$)

$$\delta_X(\lambda) = \frac{\Delta r^X_\lambda}{\sigma_X} = \frac{r^X_\lambda - r^X_{\mathrm{sp}}}{\sigma_X}, \tag{3.12}$$

where r^X_λ is the steady-state firing rate of $\mathcal{S}^X$ during the stimulus and r^X_{sp} is the spontaneous firing rate of $\mathcal{S}^X$. Here and in the following expressions, the actual time course of the firing-rate response is neglected and replaced with instantaneous jumps to and from the steady-state value during stimulation. Furthermore, it is assumed that σ_X, the standard deviation of $R^X_\lambda(t)$, does not change significantly during stimulation. With the further simplification that the autocorrelation time of the readout activity is $\tau_c \approx \tau_f$, the detection rates for the θ_+ detector can be approximated as the probability that one of $n = T/\tau_f$ draws of a Gaussian random variable

exceeds θ_+. In the detection time window $(0, T_w)$, $n_s = T_s/\tau_f$ of these draws are influenced by the stimulation and $n - n_s$ are not. Therefore, the correct detection rate for θ_+ can be estimated as

$$\mathcal{CD}^+_\lambda(\theta_+) \approx 1 - p_+^{n_s}\Big(\theta_+, \delta_X(\lambda)\Big) p_+^{n-n_s}\Big(\theta_+, 0\Big). \tag{3.13}$$

where

$$\begin{aligned} p_+\Big(\theta_+, \delta_X(\lambda)\Big) &= \int_{-\infty}^{\theta_+} \mathrm{d}a\, \mathcal{N}(a, \Delta r^X_\lambda, \sigma_X) \\ &= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\theta_+ - \Delta r^X_\lambda}{\sqrt{2}\sigma_X}\right)\right] \\ &= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\theta_+}{\sqrt{2}\sigma_X} - \frac{\delta_X(\lambda)}{\sqrt{2}}\right)\right]. \end{aligned} \tag{3.14}$$

Analogously, the false positive rate for θ_+ is

$$\mathcal{FP}^+_\lambda(\theta_+) \approx 1 - p_+^n\Big(\theta_+, 0\Big). \tag{3.15}$$

The last equation can be combined with eq. (3.11) and solved for the threshold-to-noise ratio that gives the false positive rate of 0.25:

$$\frac{\bar{\theta}_+}{\sqrt{2}\sigma_X} = \mathrm{erf}^{-1}\left(2\left(\frac{3}{4}\right)^{\tau_f/T_w} - 1\right). \tag{3.16}$$

The effect size is finally obtained by using the last equation with

$$\bar{\mathcal{Y}}^+(\lambda) = p_+\Big(\bar{\theta}_+, 0\Big)^{n-n_s}\left[p_+\Big(\bar{\theta}_+, 0\Big)^{n_s} - p_+\Big(\bar{\theta}_+, \delta_X(\lambda)\Big)^{n_s}\right]. \tag{3.17}$$

The effect size for the θ_- detector is the same where p_+ is replaced by $p_- = 1 - p_+$.

The sign of the SNR δ_X indicates whether the average deviation from the spontaneous state is in the positive or negative direction. The double-barrier detector of chapter 2 always returns a positive effect size regardless of the sign of the SNR, although the effect size measured from simulations can turn out negative because of finite-size fluctuations. On the contrary, the two single-barrier detectors considered here produce a signed effect size, plotted in fig. 3.3 as a function of the SNR. The θ_+ detector yields a positive $\bar{\mathcal{Y}}$ for a positive δ_X and a negative $\bar{\mathcal{Y}}$ for a negative δ_X (fig. 3.3A). In other words, the effect size is a monotonically increasing function of the SNR. However, the curve is not symmetrical around $\delta_X = 0$, and for negative SNR the effect size quickly saturates at $\bar{\mathcal{Y}}^+(-\infty) = \frac{3}{4}\Big(1 - \frac{3}{4}^{-n_s/n}\Big)$. For the θ_- detector, the situation is reversed: the effect size is a monotonically decreasing function of the SNR and the sign

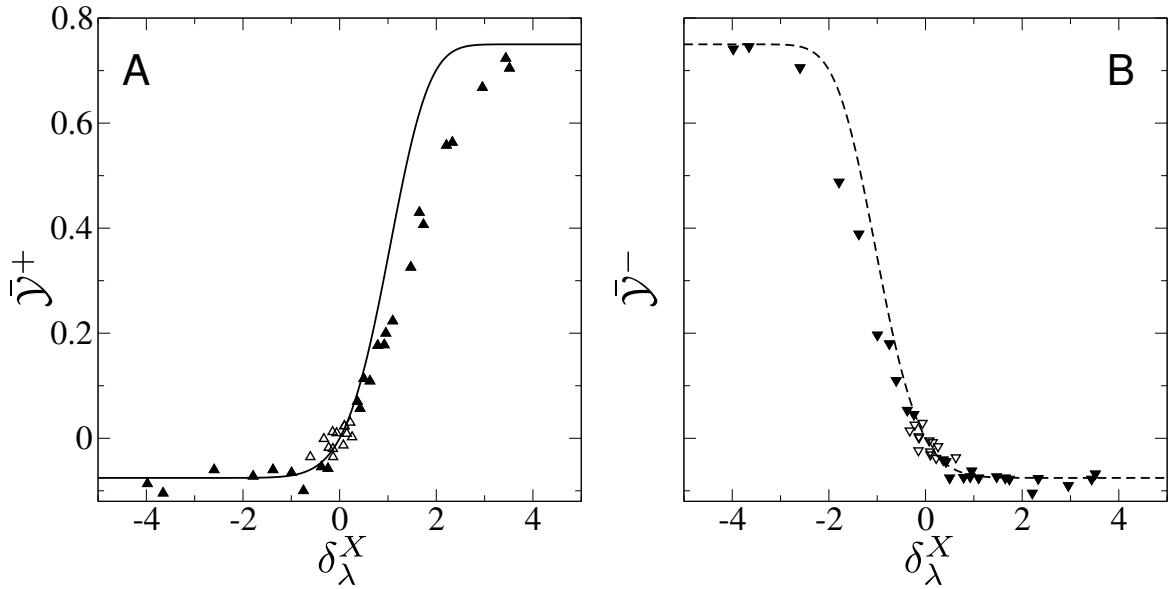

Figure 3.3. – **Effect size is a monotonic function of the signal-to-noise ratio**. **A**: Effect size as a function of the signal-to-noise ratio (SNR) for the θ_+ detector. **B**: Same for θ_- detector. Data points are taken from various dataset presented in the following sections.

of $\bar{\mathcal{Y}}$ is opposite to that of δ_X (fig. 3.3B). As a matter of fact, $\bar{\mathcal{Y}}^-(\delta_X) = -\bar{\mathcal{Y}}^+(-\delta_X)$. Taking these considerations into account, the SNR will be used to interpret and compare the qualitative behavior of the effect size in the two readout schemes. As in the previous chapter, the numerator of the SNR will be computed fully analytically, while the readout variance σ_X^2 will be mostly measured from numerical simulations.

In the next three sections, the effectiveness of the readout schemes A and B will be compared for three possible architectures of the RN, in order of increasing complexity: i) a population of neurons receiving only feed-forward input and no recurrent connections, ii) a population of neurons receiving both feed-forward excitation and local recurrent inhibition, and iii) a fully recurrent E-I network receiving feed-forward input from the BCN.

3.2. Purely feed-forward readout

In the simplest scenario, the readout is performed by a population of LIF neurons receiving input from the BCN and additional shot-noise, but no recurrent connections (fig. 3.4). In this case, the RN and the readout population $\mathcal{S}^B$ coincide. Each neuron in $\mathcal{S}^B$ obeys

$$\tau_m \frac{\mathrm{d}v_k}{\mathrm{d}t} = -v_k(t) + R_m[I_{\mathrm{ext},k}(t) + I_{\mathrm{FF},k}(t)] \tag{3.18}$$

with the usual fire-and-reset rule. The input from the BCN $I_{\mathrm{FF},k}(t)$ is

$$I_{\mathrm{FF},k}(t) = \frac{\tau_m}{R_m} \sum_{j \in \mathcal{Q}^\lambda(k)} J^{\mathrm{FF}}_{kj} x_j(t - D^{\mathrm{FF}}_{kj}), \tag{3.19}$$

where the size of $\mathcal{Q}_\lambda(k)$ is $\hat{C} = 4000$. The term modeling the input from other areas, $I_{\mathrm{ext},k}(t)$, is the same as for neurons in the BCN and is described by eq. (3.3). Hence, each neuron in the RN receives the same amount of excitatory spikes per unit time as neurons in the BCN, but no inhibitory spikes. To compensate for the missing inhibition, the constant input term is set here to $R_m I_0 = -18.0\,\mathrm{mV}$. In this way, the spontaneous firing rate of $\mathcal{S}^B$, r^B_{sp}, is approximately equal to the spontaneous firing rate of the BCN, r_{sp}, i.e. $r^B_{\mathrm{sp}} \approx r_{\mathrm{sp}} \approx 2\,\mathrm{Hz}$.

Although the mean firing rates of the BCN and the RN are similar, the spontaneous activities of the two networks look very different, as it clearly results from a comparison of the two raster plots in fig. 3.5: the firing pattern of the BCN is asynchronous, whereas the spiking activity of the RN displays synchronized fluctuations. As a consequence, the two filtered readout activities, $R^A_\lambda(t)$ and $R^B_\lambda(t)$, fluctuate around the same value but the amplitude of the fluctuations is quite different: for $R^A_\lambda(t)$ it is quite small ($\sigma_A \approx 0.09\,\mathrm{Hz}$) whereas for $R^B_\lambda(t)$ it is larger by more than one order of magnitude ($\sigma_B \approx 1.6\,\mathrm{Hz}$). These synchronized fluctuations in the activity of the RN are caused by the correlations in the input noise (Kruscha and Lindner, 2015, 2016, derived analytical approaches for the activity of a population of uncoupled LIF neurons driven by a common white noise).

To understand why the fluctuations in the readout activity are so strong, one can start from the decomposition of the readout variance into two terms derived in eq. (2.65):

$$\sigma^2_B \approx \frac{S^{\mathcal{E}}_{xx}(0)}{\sqrt{\pi}\tau_f N_B} + \frac{S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0)}{\sqrt{\pi}\tau_f,}. \tag{3.20}$$

where $S^{\mathcal{E}}_{xx}(0)$ is the low-frequency limit of the single-spike-train average power-spectrum and $S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0)$ is the low-frequency limit of the average cross-spectrum between neurons in $\mathcal{S}^B$. The magnitude of the two terms in eq. (3.20) is quite different, namely $S^{\mathcal{E}}_{x_1x_2}(0) \approx 0.5\,\mathrm{Hz}$ and $S^{\mathcal{E}}_{xx}(0)/N_B \approx r^B_{\mathrm{sp}}/N_B \approx 2 \cdot 10^{-4}\,\mathrm{Hz}$. Therefore, the first term can be neglected. Because neurons

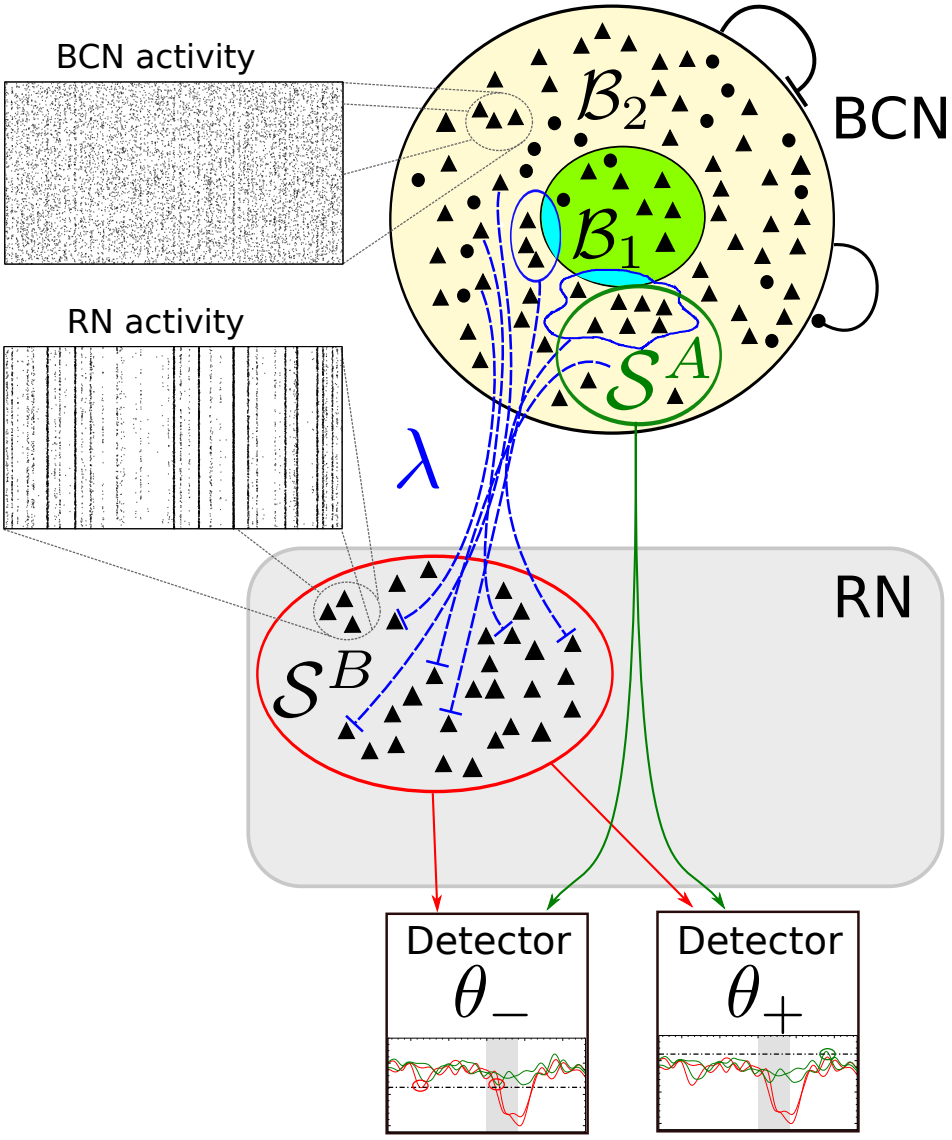

Figure 3.4. – **Purely feed-forward architecture for the readout network**. Here, the readout network consists only of the population $\mathcal{S}^B$, which has no output connections in the model. Each neuron in $\mathcal{S}^B$ receives $\hat{C} = C_E = 4000$ input connections from the barrel cortex network (BCN), depicted as blue dashed lines. These connections can be biased towards $\mathcal{B}_1$ and λ represents here the probability that a feed-forward connection to $\mathcal{S}^B$ originates from $\mathcal{B}_1$. The two detectors receive as input both the activity of the readout population $\mathcal{S}^B$ (red arrows) and directly from a readout subset of the BCN (green arrows). The two insets show the raster plot of the spontaneous activity of 4000 randomly selected neurons in the BCN and in $\mathcal{S}^B$.

in $\mathcal{S}^B$ are not recurrently connected, cross-correlations between them are caused only by cross-correlations in their inputs. As a consequence, in the linear approximation, the cross-spectrum between their spike trains is proportional to the cross-spectrum between the input to the two neurons (see section A.1),

$$S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0) \approx |\chi(0)|^2 \, S^{\mathcal{E}}_{\eta_1\eta_2}(0) \approx \left|\frac{\mathrm{d}\phi_{\mathrm{sn}}}{\mathrm{d}\mu}\right|^2 S^{\mathcal{E}}_{\eta_1\eta_2}(0). \tag{3.21}$$

In the last equation, the derivative of ϕ_{sn} is taken with respect to the mean input $\mu = R_m I_0$ (see section 1.4.2) and $S^{\mathcal{E}}_{\eta_1\eta_2} = \langle \tilde{\eta}_1 \tilde{\eta}_2^* \rangle$, where η_k is the input to a neuron in $\mathcal{S}^B$. The input cross-spectrum can be calculated by taking the explicit expression for the Fourier transform of η_k

$$\tilde{\eta}_k = \tau_m \sum_{j \in \mathcal{Q}^\lambda(k)} J_{kj} e^{2\pi i D_{kj} f} \tilde{x}_j, \tag{3.22}$$

then multiplying it by its complex conjugate and finally averaging over each pair (the calculation is analogous to the one in section A.3 with the simplification $g = 0$). Inserting the result into eq. (3.21) yields

$$\begin{aligned} S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0) &\approx \left|\frac{\mathrm{d}\phi_{\mathrm{sn}}}{\mathrm{d}\mu}\right|^2 \tau_m^2 J^2 \Big(\lambda_c \hat{C} S^E_{xx}(0) + (\hat{C}^2 - \lambda_c \hat{C}) S^{EE}_{x_1x_2}(0)\Big) \\ &\approx \alpha^2 \left(\frac{\lambda_c S^E_{xx}(0)}{\hat{C}} + S^{EE}_{x_1x_2}(0)\right). \end{aligned} \tag{3.23}$$

In eq. (3.23), $\lambda_c \hat{C}$ is the average number of neurons in the BCN providing input to both neurons, i.e. the average number of inputs shared by any couple of neurons in the RN. One finds that $\lambda_c = \lambda^2 + (1-\lambda)^2/(1-\lambda_0)\lambda_0 \approx \lambda^2 + (1-\lambda)\lambda_0$. These common inputs produce a term proportional to the low-frequency limit of the power spectrum of excitatory neurons within the BCN, $S^E_{xx}(0)$, whereas all non-diagonal terms lead to the term proportional to $S^{EE}_{x_1x_2}(0)$, the low-frequency limit of the cross spectrum between excitatory neurons within the BCN. For convenience, the symbol α has been introduced to indicate the linearization of the input-output firing rate relation:

$$\alpha = \tau_m J \hat{C} \frac{\mathrm{d}\phi_{\mathrm{sn}}}{\mathrm{d}\mu}. \tag{3.24}$$

The term proportional to the power spectrum in eq. (3.23) is small compared to the other one, except for large values of λ. Inserting eq. (3.23) without the term $\lambda_c S^E_{xx}(0)/\hat{C}$ into eq. (3.20) yields

$$\sigma_B^2 \approx \frac{S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0)}{\sqrt{\pi}\tau_f} \approx \frac{\alpha^2 S^{EE}_{x_1x_2}(0)}{\sqrt{\pi}\tau_f}. \tag{3.25}$$

The variance of $R_\lambda^A(t)$ can be decomposed in the same way as σ_B^2:

$$\sigma_A^2 \approx \frac{S_{xx}^E(0)}{\sqrt{\pi}\tau_f N_B} + \frac{S_{x_1x_2}^{EE}(0)}{\sqrt{\pi}\tau_f,}. \tag{3.26}$$

Substituting the last equation into eq. (3.25) leads to

$$\sigma_B^2 \approx \frac{\alpha^2 S_{x_1x_2}^{EE}(0)}{\sqrt{\pi}\tau_f} \approx \alpha^2 \left[\sigma_A^2 - \frac{S_{xx}^E(0)}{N_A\sqrt{\pi}\tau_f}\right]. \tag{3.27}$$

Finally, if the second term related to the power spectrum in the last equation is also neglected, a simple proportionality between the variances of $R_\lambda^A(t)$ and $R_\lambda^B(t)$ results:

$$\sigma_B^2 \approx \alpha^2\sigma_A^2. \tag{3.28}$$

Although some of these approximations are not very precise, the measured ratio of the two standard deviations $\sigma_B/\sigma_A \approx 18$ is not too far from the predicted value of $\alpha \approx 16$.

A linear-response analysis can also be applied to the firing-rate response to the stimulation. As already stated in section 3.1, the actual time-course of the firing rate response will be ignored in this chapter and only steady-state values will be considered. During the stimulus, the firing-rate deviation of each population $\mathcal{B}_k$ $(k = 0, 1, 2)$ is indicated by Δr_k. Because λ indicates the fraction of neurons in $\mathcal{S}^A$ chosen from $\mathcal{B}_1$ and $(1 - \lambda)$ is the fraction of neurons from $\mathcal{B}_2$, the deviation from the spontaneous state of the firing rate of $\mathcal{S}^A$ is

$$\Delta r_\lambda^A = \lambda\Delta r_1 + (1 - \lambda)\Delta r_2, \tag{3.29}$$

where Δr_1 and Δr_2 can be calculated from eqs. (2.14) and (2.15), as explained on p. 46 onwards. In the linear response approximation, the firing-rate response of neurons in $\mathcal{S}^B$ is

$$\Delta r_\lambda^B \approx \frac{\mathrm{d}r_{\mathrm{sp}}^B}{\mathrm{d}\mu}\Delta\mu \approx \frac{\mathrm{d}\phi_{\mathrm{sn}}}{\mathrm{d}\mu}\tau_m J(\lambda\hat{C}\Delta r_1 + (1 - \lambda)\hat{C}\Delta r_2) = \alpha\Delta r_\lambda^A, \tag{3.30}$$

where $\Delta\mu$ indicates the deviation of the mean input to each neuron from the spontaneous state. This last equation states that in the linear-response approximation the numerators of the two SNRs are also proportional to each other. Combining the two results leads to the prediction that $\delta_A(\lambda)$ and $\delta_B(\lambda)$ should be approximately equal

$$\delta_B(\lambda) = \frac{\Delta r_\lambda^B}{\sigma_B} \approx \frac{\alpha\Delta r_\lambda^A}{\alpha\sigma_A} = \delta_A(\lambda). \tag{3.31}$$

These considerations suggest that the two readout schemes should detect the single-cell stimulation with similar efficacy. The effect size obtained by the two detection schemes is plotted in

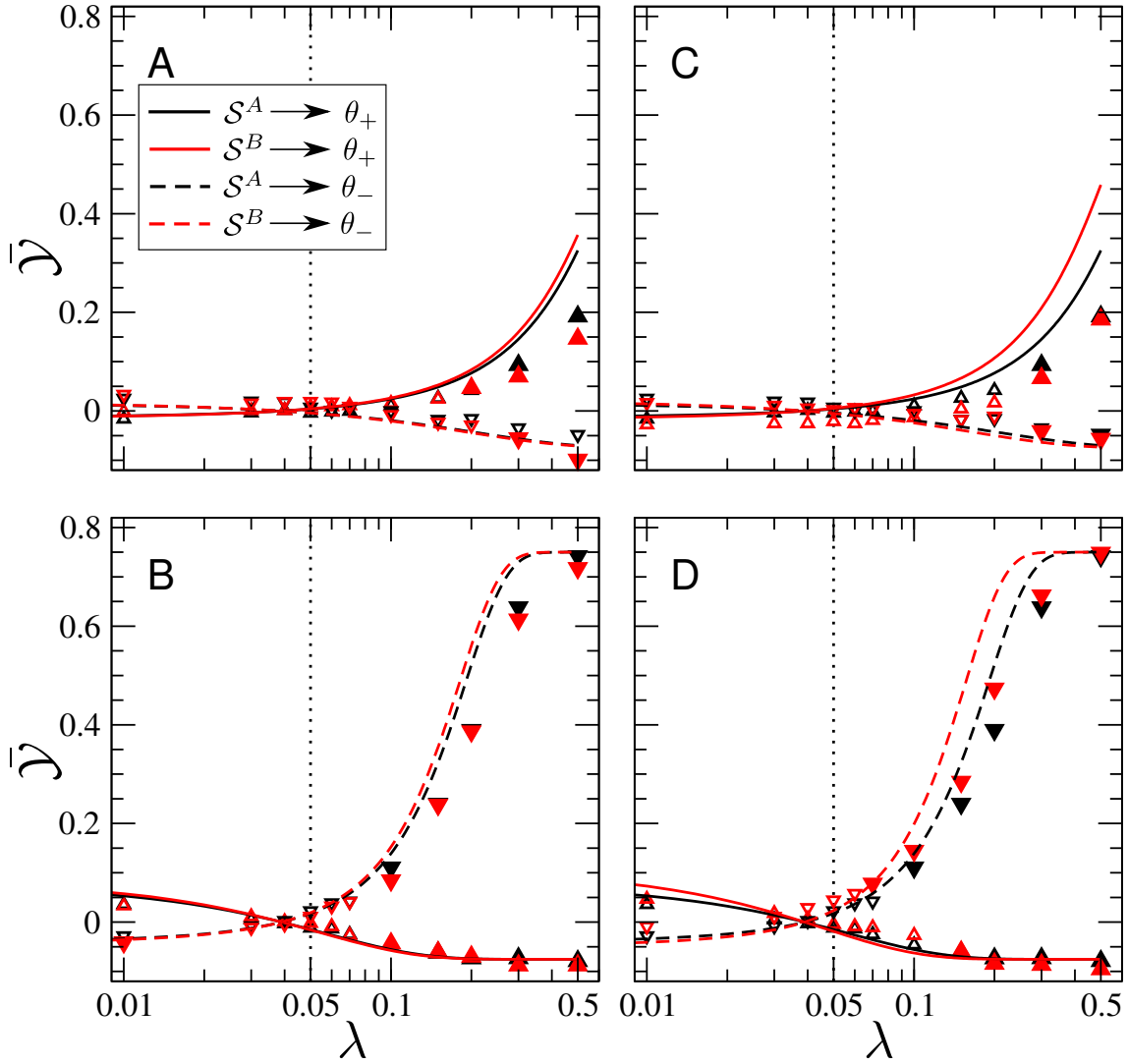

Figure 3.5. – **Purely feed-forward architecture does not improve detectability.** In all panels, results for the readout setup B are depicted in red, while results for the readout setup A are plotted in black. Simulation results for the θ_+ detector are indicated by upward pointing triangles while data points for the θ_- are shown as downward pointing triangles. Filled symbols represent data points significantly different than zero (p-value<0.05). Continuous and dashed lines represent the theory for θ_+ and θ_- detector, respectively (see inset). The four panels refer to four different cases. **A**: Detectability of an excitatory $\mathcal{B}_0$ when the spontaneous firing rate of readout and barrel cortex network match $r^B_{\rm sp} \approx r^A_{\rm sp} \approx 2\,$Hz. **B**: Same as **A** but in the case that $\mathcal{B}_0$ is inhibitory. **C**: Detectability of excitatory $\mathcal{B}_0$ when the readout spontaneous firing rate is higher $r^B_{\rm sp} \approx 15\,$Hz. **D**: Same as **C** but in the case that $\mathcal{B}_0$ is inhibitory. Parameters as in table 3.1.

fig. 3.5 as a function of the bias λ (the vertical dotted line marks $\lambda = \lambda_0$, i.e. no readout bias). Black symbols and lines refer to the readout scheme A, while red symbols and lines pertain to the readout setup B. To begin with, consider the case of excitatory $\mathcal{B}_0$, shown Figure 3.5A. Focusing first on the results for the θ_+ detector (upward pointing triangles for simulations and solid lines for theory), it can be seen that a statistically significant detection (closed symbols indicate $p < 0.05$) is observed only for large values of the bias and that the readout scheme A performs slightly better than scheme B. Only when $\lambda \leq 0.3$ data obtained from the lower barrier detector θ_- (downward pointing triangles for simulations and dashed lines for theory) are significantly different from zero, but the effect size is negative (i.e. the detector reacts less frequently to a stimulus than chance). When the stimulated cell is inhibitory (fig. 3.5B), the roles of the two detectors are reversed, and the effect size is generally stronger and quite similar for the two readout schemes. Here, the minimal bias required for a significant detection is $\lambda \approx 0.1$. The reason why inhibitory cells are easier to detect is explained in more detail in section 2.4, but it is essentially the larger average weight of inhibitory connections.

These results confirm the prediction of eq. (3.31) that the effect size for the two readout schemes is quite similar because the effect of the input-output transformation performed by the LIF model is approximately linear. However, for different parameters the situation may change. For instance, instead of requiring the firing rate of the RN to be equal to that of the BCN, the mean input to the readout neurons can be increased so that the spontaneous firing rate of $\mathcal{S}^B$ rises to $r^B_{\rm sp} \approx 15\,{\rm Hz}$. In this case, the effect size resulting from scheme B is slightly larger than in the previous case (nothing changes for the scheme A) both when $\mathcal{B}_0$ is excitatory (fig. 3.5C) when $\mathcal{B}_0$ is inhibitory $\mathcal{B}_0$ (fig. 3.5D). However, the difference is modest and the two readout schemes are still essentially equivalent.

To summarize the results of this section, using a population of cells integrating input from the BCN with no recurrent connection does not confer a clear benefit to the readout scheme B over the readout scheme A. The reason is that the input-output relation of the LIF model amplifies both signal and noise in a similar way. If this transformation were performed by a linear system, the SNR would not change under any circumstance. However, because the LIF neuron model is nonlinear, there is a chance that tinkering around with parameters of the RN can improve the SNR for the readout scheme B. For instance, increasing the output firing rate of the readout did enhance the SNR for the readout scheme setup B. However, the improvement was small, and firing rates in the cortex are typically low. Another way to improve the SNR would be to enlarge N_B, the size of $\mathcal{S}_B$, thus decreasing the first term in eq. (3.20) and, hence, the readout variance σ^2_B. However, even in the limit $N_B \to \infty$, the enhancement would be limited by the fluctuations due to cross-correlations, which are not averaged out. Therefore, even if biological constraints are disregarded, the architecture considered in this section does not seem to have the potential to improve the effectiveness of the readout scheme B in a significant way.

3.3. Readout with recurrent inhibition

Modeling the readout network as a collection of LIF neurons with no recurrent connections is the simplest option but ignores a fundamental component of cortical circuits, i.e. inhibitory neurons. In this section, the RN model is extended to include a second population of inhibitory neurons $\mathcal{I}$ in addition to the population $\mathcal{S}^B$.

As for the BCN, the ratio of excitatory to inhibitory neurons for the RN is four to one. Hence, the size of the population of inhibitory neurons $\mathcal{I}$ is $N_\mathcal{I} = \gamma N_B = 2500$. Because cortical feed-forward connections do not target specifically excitatory neurons (Feldmeyer et al., 2013), both $\mathcal{S}^B$ and $\mathcal{I}$ receive the same amount of feed-forward excitatory inputs from the BCN. Local inhibitory connections originating from $\mathcal{I}$ target both $\mathcal{S}^B$ and $\mathcal{I}$. The model is depicted in fig. 3.6. All neurons within the RN evolve according to the same equation (with fire-and-reset rule):

$$\tau_m \dot{v}_k(t) = -v_k(t) + R_m[I_{\text{ext},k}(t) + I_{\text{rec},k}(t) + I_{\text{FF},k}(t)]. \tag{3.32}$$

In eq. (3.32), the external input term $I_{\text{ext},k}(t)$ is the same as in the two previous sections. The recurrent input is purely inhibitory

$$I_{\text{rec},k}(t) = \frac{\tau_m}{R_m}\left(-g \sum_{\ell \in \mathcal{L}_i(k)} J^r_{k\ell} x_\ell(t - D^r_{k\ell})\right), \tag{3.33}$$

where $\mathcal{L}_i(k)$ are sets of $C_\mathcal{I} = 1000$ neurons selected at random from $\mathcal{I}$. The feed-forward input to neuron k in the RN is

$$I_{\text{FF},k}(t) = \frac{\tau_m}{R_m} \sum_{j \in \mathcal{Q}^\lambda_e(k)} J^{\text{FF}}_{kj} x_j(t - D^{\text{FF}}_{kj}), \tag{3.34}$$

where $\lambda = \lambda_e$ if neuron k is in $\mathcal{S}^B$ and $\lambda = \lambda_i$ if it is in $\mathcal{I}$. In other words, the bias of connections to excitatory and inhibitory neurons in the RN is controlled by two separate parameters. To avoid confusion, from now on the bias parameter for the readout scheme A is indicated with λ_A.

Neurons in the RN receive the same total amount of excitatory and inhibitory spikes as neurons within the BCN. Therefore, it is not surprising that both populations within the RN ($\mathcal{S}^B$ and $\mathcal{I}$) have almost the same spontaneous firing rate as neurons within the BCN, i.e. $r^B_{\text{sp}} \approx r^\mathcal{I}_{\text{sp}} \approx r_{\text{sp}} \approx 2\,\text{Hz}$. Furthermore, the spiking activity of the RN is asynchronous like that of the BCN (raster plots in fig. 3.6).

As mentioned above, the readout bias of the excitatory population $\mathcal{S}^B$ and of the inhibitory population $\mathcal{I}$ are governed by two distinct parameters. To begin with, one can first consider the scenario in which the learning process involves only the feed-forward connections to $\mathcal{S}^B$. To this end, λ_e is varied while the bias to inhibitory neurons in the RN is fixed and equal to its

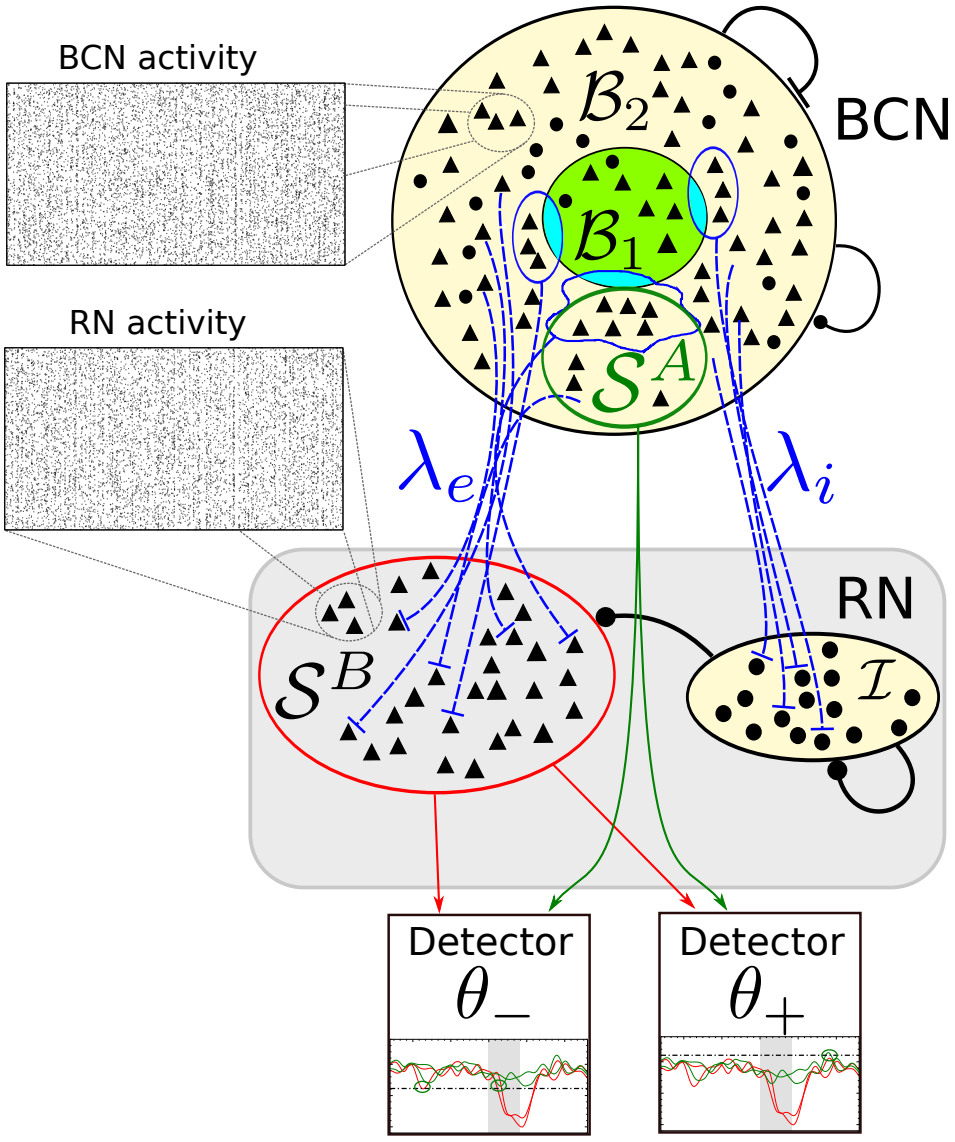

Figure 3.6. – **Architecture with local inhibition for the readout network**. Here, the readout network consists only two populations: the readout population $\mathcal{S}^B$, which has no output connections as in the previous case, and a population of inhibitory interneurons $\mathcal{I}$, which are recurrently connected to each other and provide feed forward inhibition to $\mathcal{S}^B$. Each neuron in the RN receives $\hat{C} = C_E = 4000$ input connections from the barrel cortex network (BCN), depicted as blue dashed lines. These connections can be biased towards $\mathcal{B}_1$, where λ_e represents the bias in the connections from the BCN to $\mathcal{S}^B$, and λ_e represents the bias in the connections from the BCN to $\mathcal{I}$. As in the previous configuration, the two detectors receive as input both the activity of the readout population $\mathcal{S}^B$ (red arrows) and directly from a readout subset of the BCN (green arrows). The two insets show the raster plot of the spontaneous activity of 4000 randomly selected neurons in the BCN and in $\mathcal{S}^B$.

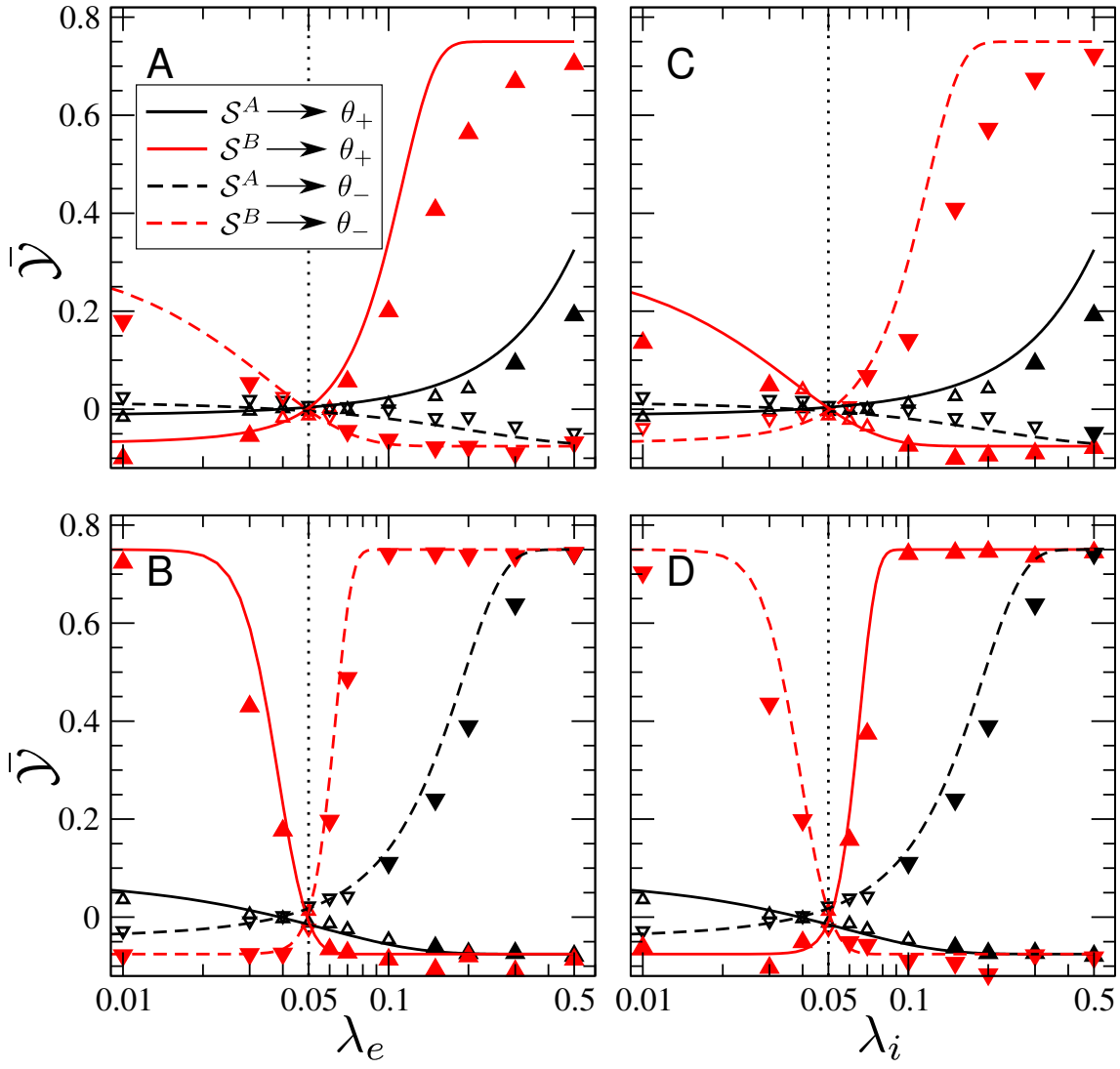

Figure 3.7. – **Local inhibition greatly enhances the ability to detect the single-cell stimulation.** In all panels, meaning of colors and symbols is the same as in fig. 3.5 (black for readout scheme A, red for readout scheme B, solid lines and upward pointing triangles for θ_+ detector, dashed lines and downward pointing triangles for θ_- detector). **A**: Detectability for excitatory $\mathcal{B}_0$ when varying bias λ_e from the BCN to $\mathcal{S}^B$, where the bias to $\mathcal{I}$ is fixed $\lambda_i = \lambda_0$; the bias of the readout $\mathcal{S}^A$ is $\lambda_A = \lambda_e$. **B**: same as in **A** but for inhibitory $\mathcal{B}_0$. **C**: Detectability for excitatory $\mathcal{B}_0$ when varying bias λ_i from the BCN to $\mathcal{I}$, where the bias to $\mathcal{S}^B$ is fixed $\lambda_e = \lambda_0$; the bias of the readout $\mathcal{S}^A$ is $\lambda_A = \lambda_i$. **D**: same as in **C** but for inhibitory $\mathcal{B}_0$.

natural value $\lambda_i = \lambda_0$. To compare the two readout schemes, the bias of the readout set $\mathcal{S}^A$ is taken equal to the bias to $\mathcal{S}^B$, i.e. $\lambda_A = \lambda_e$. Figures 3.7A and 3.7B show the effect size as a function of λ_e when the stimulated cell is excitatory and inhibitory, respectively. The meaning of colors and symbols is the same as in fig. 3.5 (black for readout scheme A, red for readout scheme B, solid lines and upward pointing triangles for θ_+ detector, dashed lines and downward pointing triangles for θ_- detector). The difference between the effect size measured by the two readout schemes is here quite large. When the stimulated cell $\mathcal{B}_0$ is excitatory (fig. 3.7A) the readout scheme B produces almost everywhere a significant effect (red closed triangles) except in the case that λ_e is in the vicinity of the unbiased case i.e. $\lambda_e \approx \lambda_0$. When $\mathcal{B}_0$ is inhibitory (fig. 3.7B), the effect is even larger and there is only one data point for which the detection is *not* significant, that is $\lambda_e = \lambda_0$. The upper-boundary detector θ_+ (upward pointing triangles and solid lines) returns a large positive effect size when $\mathcal{B}_0$ is excitatory and $\lambda_e > \lambda_0$ or when $\mathcal{B}_0$ is inhibitory and $\lambda_e < \lambda_0$, the cases in which the firing-rate response of $\mathcal{S}^B$ is in the positive direction. The converse holds for the lower-boundary detector θ_-, that yields a positive effect size in the two complementary cases (inhibitory $\mathcal{B}_0$ and $\lambda_e > \lambda_0$ and excitatory $\mathcal{B}_0$ $\lambda_e < \lambda_0$), which correspond to an average decrease in the firing rate of $\mathcal{S}^B$.

The scenario in which only the feed-forward connections to $\mathcal{I}$ are rewired during the learning phase of the experiment is considered in the two panels on the right side of fig. 3.7. Here, the bias parameter λ_i is varied while the connections to $\mathcal{S}^B$ are left unbiased, i.e. $\lambda_e = \lambda_0$. Again, to compare the two readout schemes, results from scheme A are plotted with the condition $\lambda_A = \lambda_i$ (and are the same data and theory as those shown in panels A and B). Both when the stimulated cell is excitatory (fig. 3.7C) and when it is inhibitory (fig. 3.7D), the plot looks almost like a copy of the corresponding case of the previous scenario, provided that the role of the two detectors in the readout scheme B is interchanged.

In all cases shown in fig. 3.7, it is evident that the theory captures the qualitative picture rather well. Hence, the linear-response approach underlying these theoretical estimates can be used to understand the reasons for the conspicuous difference between the effect resulting from the two detection schemes, which is done in the following. The firing-rate response of inhibitory interneurons within the RN is, at the linear order,

$$\Delta r^{\mathcal{I}} \approx \frac{\mathrm{d}\phi_{\mathrm{sn}}}{\mathrm{d}\mu}\left[\tau_m J\hat{C}\lambda_i \Delta r_1 + \tau_m J\hat{C}(1-\lambda_i)\Delta r_2 - \tau_m g J C_{\mathcal{I}} \Delta r^{\mathcal{I}}\right], \tag{3.35}$$

where the three terms correspond to the feed-forward mean input from $\mathcal{B}_1$, to the feed-forward mean input from $\mathcal{B}_2$, and to recurrent mean input from $\mathcal{I}$. Performing first the substitutions $\Delta r^A_{\lambda_i} = \lambda_i \Delta r_1 + (1-\lambda_i)\Delta r_2$ and $C_{\mathcal{I}} = \gamma C_E = \gamma \hat{C}$ in the last equation, and then solving it for $\Delta r^{\mathcal{I}}$ yields

$$\Delta r^{\mathcal{I}} = \alpha \frac{\Delta r^A_{\lambda_i}}{1 + g\gamma\alpha}. \tag{3.36}$$

The firing-rate response of neurons in $\mathcal{S}^B$ can be calculated analogously

$$\Delta r^B \approx \alpha(\Delta r^A_{\lambda_e} - g\gamma\Delta r^{\mathcal{I}}) = \alpha\left(\Delta r^A_{\lambda_e} - \frac{g\gamma\alpha\Delta r^A_{\lambda_i}}{1+g\gamma\alpha}\right). \tag{3.37}$$

where, in the second step eq. (3.36) was inserted. If $\gamma g\alpha \gg 1$ and $\Delta r^A_{\lambda_i}/\Delta r^A_{\lambda_e} \approx \lambda_i/\lambda_e$ (this second approximation is not very precise if λ_e, λ_i are very small), eq. (3.37) simplifies to

$$\Delta r^B \approx \alpha\Delta r^A_{\lambda_e}\left(1 - \frac{\lambda_i}{\lambda_e}\right), \tag{3.38}$$

which expresses the deviation from the spontaneous firing rate of $\mathcal{S}^B$ as a function of the bias parameters.

When analyzing the variance of the readout activity σ^2_B, one can first decompose it in the same way as in the previous section,

$$\sigma^2_B \approx \frac{S^{\mathcal{E}}_{xx}(0)}{\sqrt{\pi}\tau_f N_B} + \frac{S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0)}{\sqrt{\pi}\tau_f}, \tag{3.39}$$

and realize that, because $\mathcal{S}^B$ are not recurrently connected, the cross-spectrum $S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0)$ can be considered, in the linear-response approximation, proportional to the input cross-spectrum $S^{\mathcal{E}}_{\eta_1\eta_2}(0)$, i.e. that eq. (3.21) is still valid. However, the expression for input cross-correlations is more complicated than in the last section because here inputs are of two types (excitatory and inhibitory) and they are mutually correlated. A direct calculation, analogous to that of the previous section and of section A.3, yields

$$S^{\mathcal{E}}_{\eta_1\eta_2}(0) \approx \tau_m^2 J^2\hat{C}^2\left[\frac{\lambda_c S^E_{xx}(0) + \hat{\lambda}\gamma g^2 S^{\mathcal{I}}_{xx}(0)}{\hat{C}} + S^{EE}_{x_1x_2}(0) - 2g\gamma S^{E\mathcal{I}}_{x_1x_2}(0) + g^2\gamma^2 S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)\right], \tag{3.40}$$

where $S^E_{xx}(0)$ and $S^{\mathcal{I}}_{xx}(0)$ are the low-frequency limits of the single spike-train power spectrum of excitatory neurons in the BCN and of inhibitory neurons in $\mathcal{I}$, respectively. Furthermore, $\hat{\lambda} = C_E/N_B$ is the average fraction of shared input spike-trains from $\mathcal{I}$, and $\lambda_c \approx \lambda_e^2 + (1-\lambda_e)^2\lambda_0$ is the average fraction of shared inputs from the BCN, as in the last section. The term $S^{E\mathcal{I}}_{x_1x_2}(0)$ represents cross-correlations between excitatory neurons in the BCN and neurons in $\mathcal{I}$, and the term $S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)$ represents cross-correlations between pairs of neurons within $\mathcal{I}$. The cross-spectra $S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)$ and $S^{E\mathcal{I}}_{x_1x_2}(0)$ depend on λ_i, which is not indicated for simplicity. As discussed in appendix A (see in particular eq. A.76), recurrent connections reduce the cross-spectrum between inhibitory pairs proportionally to the connection probability. Within $\mathcal{I}$ the connection probability is rather high ($p^{\mathcal{I}}_c \approx 0.4$), which causes the term $S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)$ to be ne negative. The term $S^{E\mathcal{I}}_{x_1x_2}(0)$ is larger and positive, because excitatory neurons in the BCN drive neurons in

$\mathcal{I}$. By taking into account the respective prefactors, it is easy to see that both terms contribute negatively to the sum in eq. (3.40) and thus reduce the total input cross-correlation for low-frequencies. As discussed in the previous chapter, the linear response ansatz eq. (3.21) is not precise. However, numerical measurements reveal that $S^{\mathcal{EE}}_{x_1x_2}(0) \approx S^{EE}_{x_1x_2}(0)$. The first term in eq. (3.39) is of secondary importance, so that in the end one finds that

$$\sigma_B^2 \approx \sigma_A^2. \tag{3.41}$$

This last relation is not valid for large values of $\lambda_e \gtrsim 0.2$, for which the term $\lambda_c S^E_{xx}(0) \approx \lambda_e^2 S^E_{xx}(0)$ in eq. (3.40) becomes large.

Combining eq. (3.41) with eq. (3.37) yields the following relation between the SNRs δ_A, δ_B of the two setups

$$\frac{\delta_B}{\delta_A} \approx \frac{\Delta r^B}{\Delta r^A_{\lambda_A}} \approx \frac{\alpha}{\Delta r^A_{\lambda_A}} \left(\Delta r^A_{\lambda_e} - \frac{g\gamma\alpha \Delta r^A_{\lambda_i}}{1 + g\gamma\alpha} \right) \approx \frac{\alpha}{\Delta r^A_{\lambda_A}} \left(\Delta r^A_{\lambda_e} - \Delta r^A_{\lambda_i} \right). \tag{3.42}$$

This last result can be used to interpret the results of fig. 3.7. Considering the first scenario of learning affecting the bias to $\mathcal{S}^B$ requires setting $\lambda_A = \lambda_e$ and $\lambda_i = \lambda_0$ in eq. (3.42). With the assumption $\gamma g \alpha \gg 1$ and $\Delta r^A_{\lambda_0} / \Delta r^A_{\lambda_e} \approx \lambda_0 / \lambda_e$, one obtains

$$\frac{\delta_B}{\delta_A} \approx \alpha \left(1 - \frac{g\gamma\alpha \Delta r^A_{\lambda_0}}{(1 + g\gamma\alpha)\Delta r^A_{\lambda_e}} \right) \approx \alpha \left(1 - \frac{\lambda_0}{\lambda_e} \right). \tag{3.43}$$

The numerical value of $\mathrm{d}\phi_{\mathrm{sn}}/\mathrm{d}\mu$ (and thus of α) is here smaller than in the previous section by about one half, probably because of the larger input noise due to the inhibitory inputs, which were absent in the previous configuration of the RN. Nevertheless, the slope of the input-output linearization $\alpha \approx 8$ is still rather steep. Therefore, if $\lambda_e \neq \lambda_0$, a considerable improvement in the SNR of the readout B compared to δ_A can be expected. Indeed, fig. 3.7A and B show that the effect size is much larger for setup B whenever λ_e is sufficiently larger or smaller than λ_0, in line with eq. (3.43). However, when λ_e is smaller than λ_0, eq. (3.43) predicts that $\delta_B/\delta_A \to -\infty$. Indeed, there is a small range ($0.04 \leq \lambda_e \leq 0.05$) where the role of the two barriers for the readout scheme B and A is indeed exchanged, which is consistent with a negative ratio of the two SNRs (recall the "anti-symmetry" between the two detectors discussed in fig. 3.3). However, for smaller values of λ_e, the two detectors yield an effect size of the same sign, which means that eq. (3.43) is no longer valid when $\lambda_e < 0.04$. Going back to eq. (3.42) (again with $\lambda_A = \lambda_e$, $\lambda_i = \lambda_0$) and letting $\lambda_e \to 0$ predicts a saturation of δ_B/δ_A to

$$\frac{\delta_B}{\delta_A} \xrightarrow[\lambda_e \to 0]{} \alpha \left(1 - \frac{\Delta r_1}{\Delta r_2} \right) \lambda_0. \tag{3.44}$$

As explained in section 2.2, the ratio between Δr_1 and Δr_2 is rather large and negative regardless of the type of the stimulated cell. Therefore, the term in brackets is much larger than one, which explains why the readout scheme B yields a better effect size also for $\lambda_e \ll \lambda_0$, as seen in fig. 3.7A and 3.7B. In the other scenario, the bias of excitatory-to-inhibitory (BCN to $\mathcal{I}$) synapses, λ_i, was varied while setting $\lambda_e = \lambda_0$ and $\lambda_A = \lambda_i$. By performing these substitutions in eq. (3.42) and making the same approximation as in eq. (3.43) one finds

$$\frac{\delta_B}{\delta_A} \approx -\alpha \left(1 - \frac{\lambda_0}{\lambda_i}\right), \tag{3.45}$$

which explains why the results in fig. 3.7C and 3.7D are almost identical to those in fig. 3.7A and 3.7B, if the role of the two detectors is interchanged.

If the bias to both $\mathcal{S}^B$ and $\mathcal{I}$ is varied in parallel, i.e. if $\lambda_e = \lambda_i = \lambda_A$, eq. (3.42) predicts that the ratio δ_B/δ_A should be approximately constant and smaller than one

$$\frac{\delta_B}{\delta_A} \approx \frac{\alpha}{1 + g\gamma\alpha} \approx 1/2. \tag{3.46}$$

Equation (3.46) implies that the effect detected by the readout scheme B should be markedly smaller than the effect measured by readout scheme A. Simulation results for this case confirm that the scheme B performs always worse than the readout A both when $\mathcal{B}_0$ is excitatory fig. 3.8A) and when $\mathcal{B}_0$ is inhibitory (fig. 3.8B).

A further prediction of the theory is that changing both bias parameters in *opposite* directions should bring a large improvement in the SNR of the readout scheme B. One way of testing is idea is to set $\lambda_e = \lambda_0 + \Delta\lambda$ and $\lambda_i = \lambda_0 - \Delta\lambda = 2\lambda_0 - \lambda_e$, which limits the possible range for λ_e to $(0, 2\lambda_0)$. Figure 3.8C demonstrates that in this case the readout scheme B can significantly detect the stimulus for $\Delta\lambda$ as small as ≈ 0.01 in either direction (the x-axis reports the effect as a function of the excitatory bias λ_e, while $\lambda_i = 2\lambda_0 - \lambda_e$ and $\lambda_A = \lambda_e$) if the stimulated cell is excitatory. When $\mathcal{B}_0$ is inhibitory, an even smaller deviation from the naive state $\Delta\lambda \approx 0.005$ is sufficient to obtain a rather large effect size of more than 10% (fig. 3.8D).

In summary, feed-forward inhibition removes to a large extent input cross-correlations, which are the main source of noise in the detection. In the untrained scenario, or if the readout bias of all feed-forward connections to the RN is the same (i.e. $\lambda_e = \lambda_i$), an analogous cancellation applies also to the signal. As a consequence, the readout scheme B is less effective than the readout scheme A in detecting the stimulation. If, however, the learning process acts on the connections to the two populations in the RN in a different way and results in $\lambda_e \neq \lambda_i$, the readout scheme B can detect the single-cell stimulation with a very little change in the readout connections.

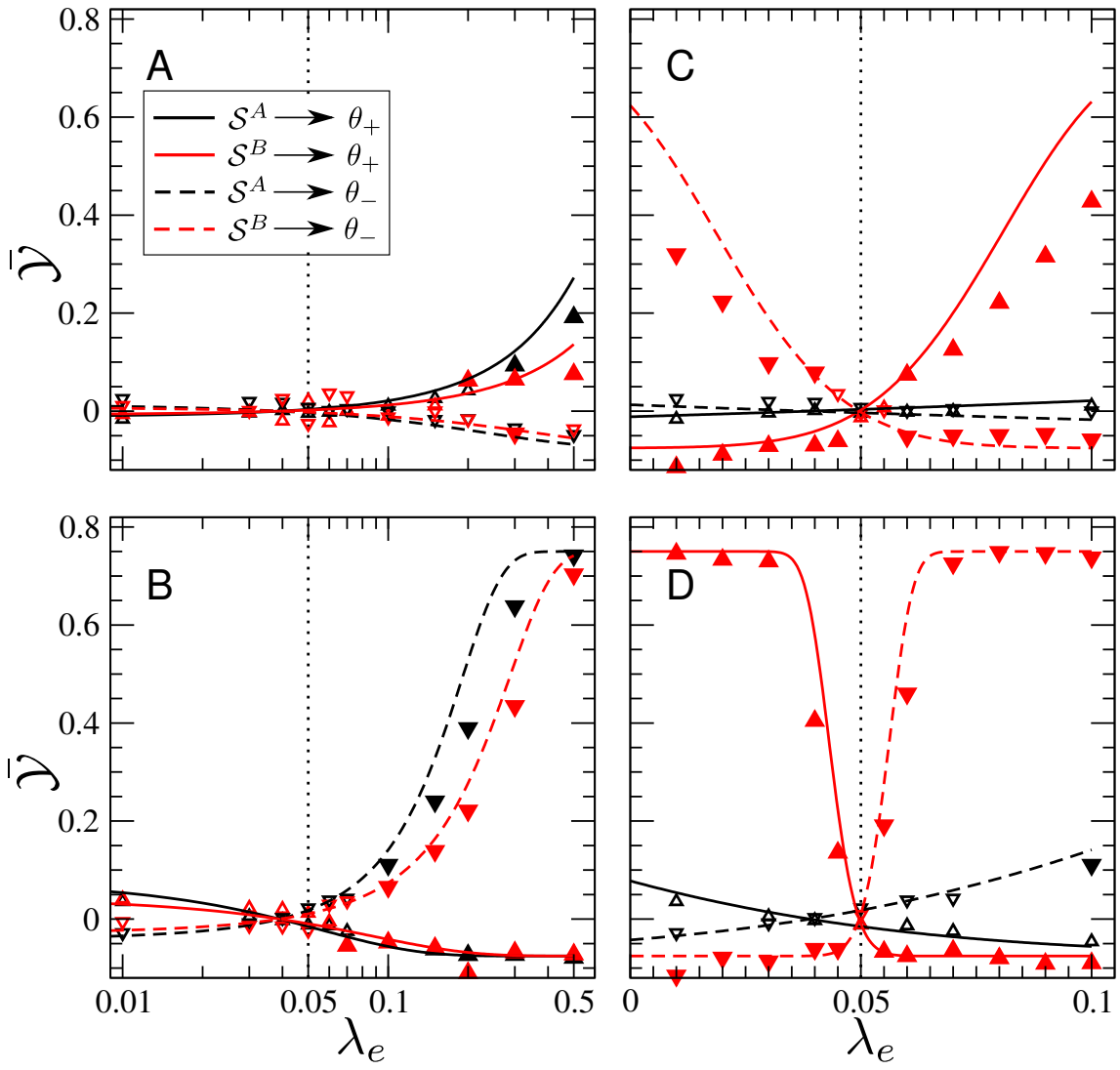

Figure 3.8. – **Changing both readout bias parameters in parallel extinguishes detectability, whereas changing them in opposite direction enhances detectability even further.** In all panels, meaning of colors and symbols is the same as in fig. 3.7. **A**: Detectability of excitatory $\mathcal{B}_0$ when varying both bias parameters in parallel, i.e. $\lambda_e = \lambda_i$; the bias of the readout $\mathcal{S}^A$ is $\lambda_A = \lambda_e = \lambda_i$. **B**: same as in **A** but for the case that $\mathcal{B}_0$ is inhibitory. **C**: Detectability of excitatory $\mathcal{B}_0$ when the bias to $\mathcal{S}^B$ and $\mathcal{I}$ are changed in opposite directions, i.e. $\lambda_e = \lambda_0 + \Delta\lambda$, $\lambda_i = \lambda_0 - \Delta\lambda$; the bias of the readout $\mathcal{S}^A$ is $\lambda_A = \lambda_e$. **D**: same as in **C** but when $\mathcal{B}_0$ is inhibitory.

3.4. Readout with a recurrent excitatory-inhibitory network

The preceding section demonstrated how adding a population of inhibitory neurons to the RN can improve both the biological realism and the efficiency of the detector circuit. However, even if input from other areas was mimicked by the external shot-noise, it is still somewhat artificial to assume that such a large fraction of the input to the RN (about 50% of all input spikes) originates from the BCN. Furthermore, recurrent connections between excitatory neurons in the RN were also completely neglected. In this section, these two issues are addressed by allowing for excitatory recurrent connections in the RN and reducing the number of inputs from the BCN to the RN. The new architecture for the RN is shown in fig. 3.9. The equation governing the evolution of each neuron's membrane potential has the same form as in the last section, i.e. eq. (3.32). The external input term is also unchanged with respect to the previous configurations. The input term from the BCN to the RN is analogous to that of the previous section, described by eq. (3.19), but the number of feed-forward inputs per neuron is here reduced to $\hat{C} = 1000$. To conserve the total number of excitatory inputs per neuron, the missing $C_{\mathcal{E}} = C_E - \hat{C} = 3000$ ones are replaced by excitatory recurrent connections from the RN. Therefore, the input term describing the recurrent input changes to

$$I_{\mathrm{rec},k}(t) = \frac{\tau_m}{R_m}\left[\sum_{j\in\mathcal{L}_e(k)} J^r_{kj} x_j(t - D^r_{kj}) - g \sum_{\ell\in\mathcal{L}_i(k)} J^r_{k\ell} x_\ell(t - D^r_{k\ell})\right], \tag{3.47}$$

where $\mathcal{L}_e(k)$ are sets of $C_{\mathcal{E}}$ neurons randomly chosen from $\mathcal{S}^B$. With this choice of parameters, only 25% of the input connections and about 12% of the total excitatory input spikes received by the RN come from the BCN.

The spontaneous activity of the RN is still asynchronous and irregular with a spontaneous firing rate of about 2 Hz, as in the previous case. However, some faint vertical stripes can be seen in the raster plot of the spiking activity (inset in fig. 3.9), which hint at stronger cross-correlations due to the recurrent excitation. These global oscillations are more evident than in the BCN because the connectivity in the RN is much denser (see appendix A for a discussion of the influence of the connectivity on cross-correlations).

For brevity, of all possibilities considered in the last section only the case that learning occurs at synapses connecting to $\mathcal{S}^B$ will be considered, which implies that $\lambda_i = \lambda_0$ and $\lambda_A = \lambda_e$. Both in the case that $\mathcal{B}_0$ is excitatory (fig. 3.10A) and when $\mathcal{B}_0$ is inhibitory (fig. 3.10B) the readout scheme B is more effective than the readout scheme A. Although the difference is less marked than in the previous section, the bias required to detect the stimulation is strongly reduced and the effect size is at least doubled for all significant data points.

The linear-response theory can be easily adapted to this case. Considerations analogous to

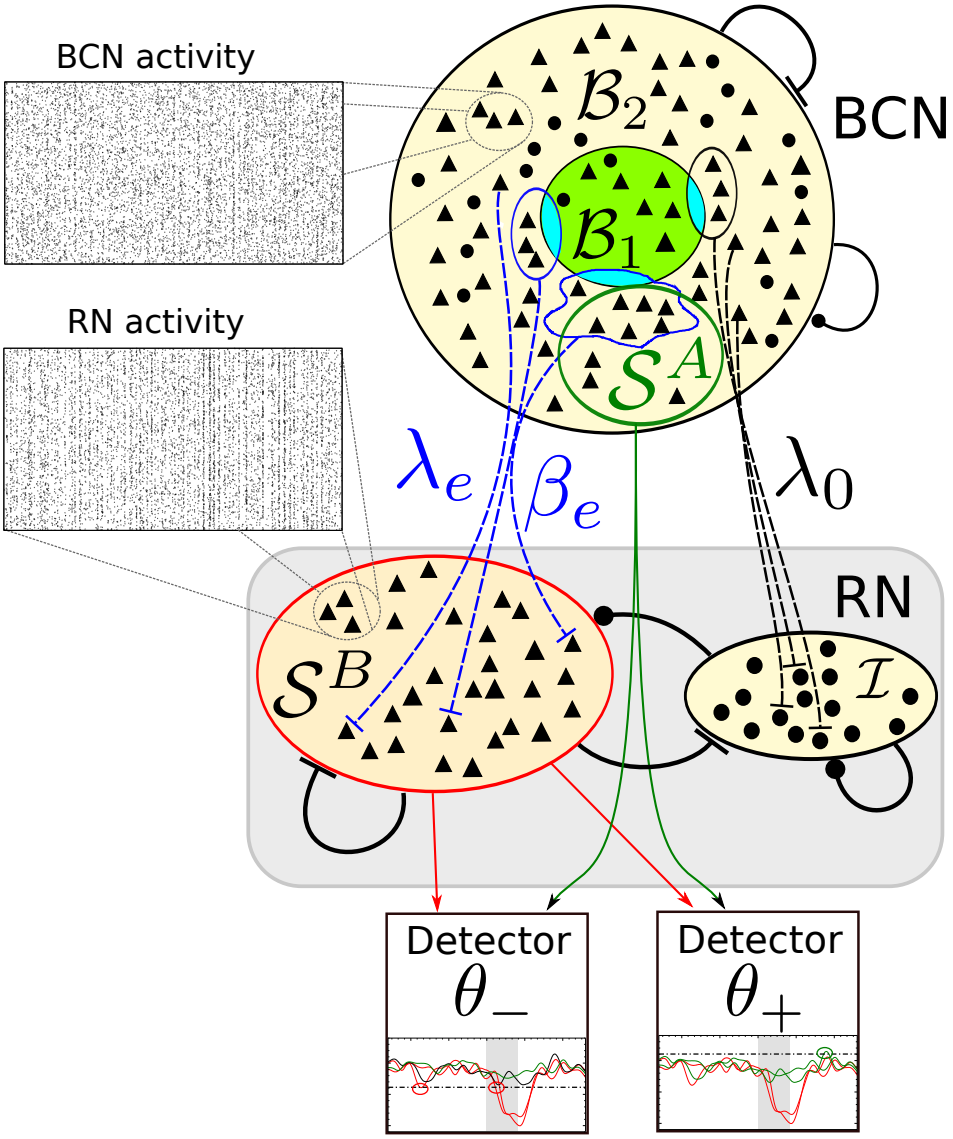

Figure 3.9. – **Fully recurrent E-I architecture for the readout network can also improve the detectability of the single-cell stimulation**. Here, $\mathcal{S}^B$ and $\mathcal{I}$ receive recurrent excitatory input from $\mathcal{S}^B$ (75% of total recurrent input) and feed-forward excitatory input from the BCN (25% of total input). Two definitions of the readout bias are used in this section. In addition to the biased connection probability λ_e, as before, the relative strength of connections from $\mathcal{B}_1$ to $\mathcal{S}^B$ is varied, indicated as β_e. The two insets show the raster plot of the spontaneous activity of 4000 randomly selected neurons in the BCN and in $\mathcal{S}^B$.

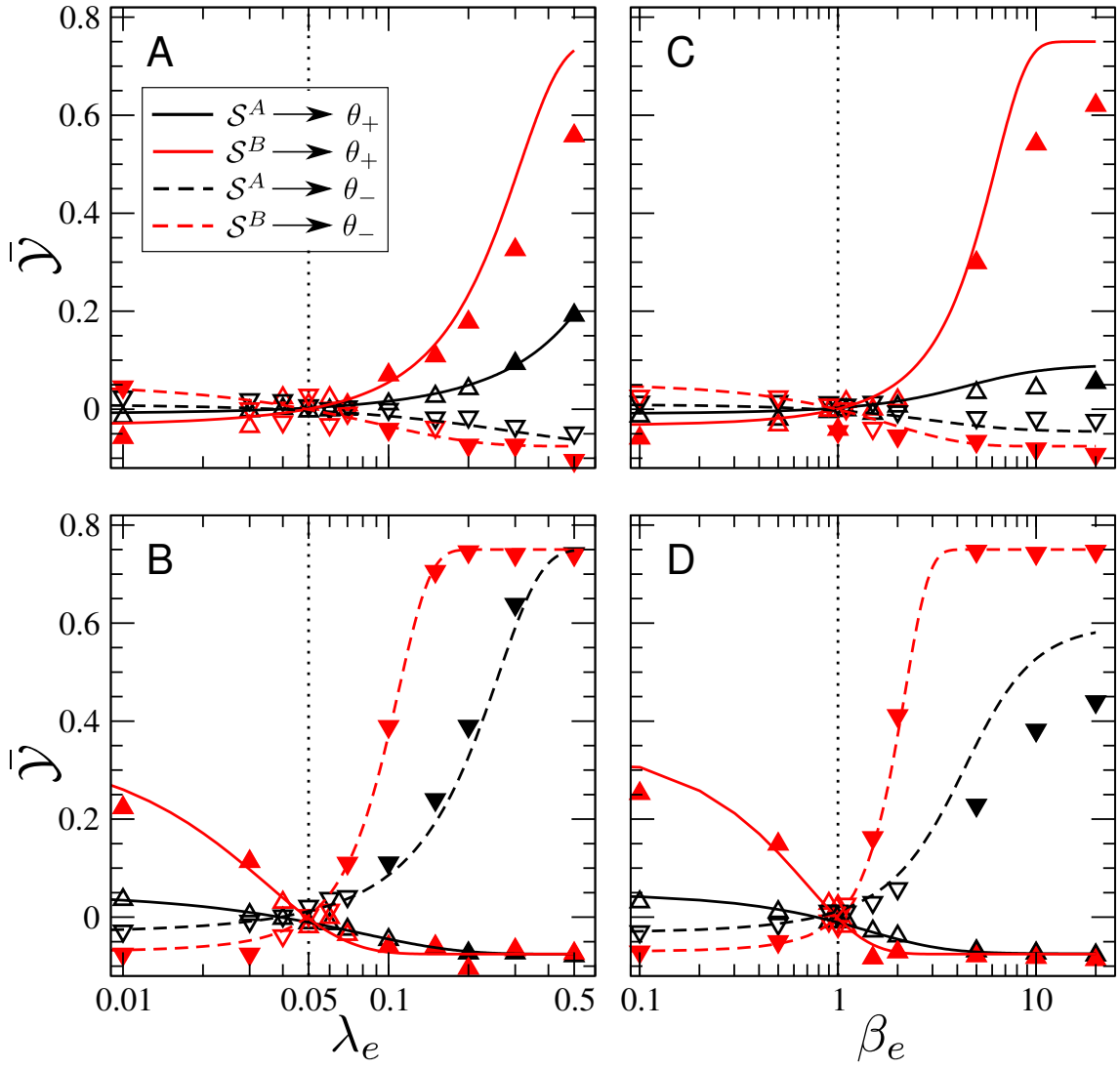

Figure 3.10. – **Fully recurrent E-I network also enhances the detectability of the single-cell stimulation**. **A**: Effect size for excitatory $\mathcal{B}_0$ when varying bias λ_e from the BCN to $\mathcal{S}^B$, where the bias to $\mathcal{I}$ is fixed $\lambda_i = \lambda_0$; the bias of the readout $\mathcal{S}^A$ is $\lambda_A = \lambda_e$. **B**: same as in **A** but for inhibitory $\mathcal{B}_0$. **C**: Effect size for excitatory $\mathcal{B}_0$ when the relative strength of connections from $\mathcal{B}_1$ to $\mathcal{S}^B$ is varied. The connection strength of other connections is adjusted to keep the total mean input constant; the bias for the other setup is implemented by a weighted sum of the spike trains within $\mathcal{S}^A$ (see text). **D**: same as in **C** but for inhibitory $\mathcal{B}_0$. In all panels, meaning of colors and symbols is the same as in the previous plots of this chapter.

those in the previous section lead to

$$\frac{\delta_B}{\delta_A} \approx \hat{\alpha}\left(1 - \frac{\lambda_0}{\lambda_e}\right). \tag{3.48}$$

The effective amplification $\hat{\alpha}$ summarizes the combined effect of feed-forward and recurrent connections (to the linear order)

$$\hat{\alpha} = \alpha\left[1 - \frac{\alpha C_{\mathcal{E}}}{\hat{C} + g\gamma\alpha(\hat{C} + C_{\mathcal{E}})}\right]^{-1}. \tag{3.49}$$

where $\alpha = \tau_m J\hat{C}\frac{\mathrm{d}\phi_{\mathrm{sn}}}{\mathrm{d}\mu}$, as in the previous sections. In this section, $\alpha \approx 2$ is smaller by roughly a factor four compared to section 3.3 because of its proportionality to $\hat{C}$. This reduction is partially compensated by the factor in square brackets in eq. (3.49), which has magnitude ≈ 1.7 and represents the effect of recurrent excitation. In the end, $\hat{\alpha}$ is still significantly larger than one, which explains why δ_B is significantly larger δ_A when λ_e is sufficiently different from λ_0.

The idea that the connection probability can be biased towards or against $\mathcal{B}_1$ supposes that some rewiring of the graph has to take place. Although formation and elimination of synapses has indeed been observed in the adult brain as a consequence of learning (Chklovskii et al., 2004), it is interesting to ascertain whether the bias can be implemented by changing the synaptic coupling instead of the connection probability. To explore this question, the readout architecture of fig. 3.9 is considered as a starting point and connections between the BCN and the readout are formed with equal probability from $\mathcal{B}_1$ and $\mathcal{B}_2$. In other words, neurons from $\mathcal{B}_1$ are chosen with probability λ_0 and neurons from $\mathcal{B}_2$ with probability $1 - \lambda_0$. However, connections originating from $\mathcal{B}_1$ are assigned a weight drawn from an exponential distribution with mean $\beta_e J$ and those from $\mathcal{B}_2$ are drawn from an exponential distribution with mean $J(1 - \beta_e\lambda_0)/(1 - \lambda_0)$. By this construction, the average coupling amplitude is J regardless of the value of β_e. The definition of $\hat{J}$ also requires β_e to be at most $1/\lambda_0$, otherwise $\hat{J}$ would become negative. Hence, the feed-forward input to each neuron in the RN reads now

$$I_{\mathrm{FF},k}(t) = \frac{\tau_m}{R_m}\left(\sum_{j\in\mathcal{Q}_e(k)\cap\mathcal{B}_1} \beta_e J^{\mathrm{FF}}_{kj} x_j(t - D^{\mathrm{FF}}_{kj}) + \sum_{i\in\mathcal{Q}_e(k)\cap\mathcal{B}_2} \frac{1 - \beta_e\lambda_0}{1 - \lambda_0} J^{\mathrm{FF}}_{ki} x_i(t - D^{\mathrm{FF}}_{ki})\right). \tag{3.50}$$

To realize the new definition of the readout bias in the readout scheme A, neurons in $\mathcal{B}_1$ and $\mathcal{B}_2$ are weighted differently by the detector. More precisely, the readout activity for the scheme A is now

$$R^A_{\beta_e}(t) = \frac{1}{N_A}\left[\beta_e \sum_{j\in\mathcal{S}^A\cap\mathcal{B}_1} x_j(t) + \frac{1 - \beta_e\lambda_0}{1 - \lambda_0} \sum_{i\in\mathcal{S}^A\cap\mathcal{B}_2} x_i(t)\right] \star F_{\tau_f}(t), \tag{3.51}$$

where the prefactor preceding the second sum guarantees that the mean of $R^A_{\beta_e}(t)$ does not

depend on β_e.

Figure 3.10C shows the effect size as a function of β_e when the stimulated cell is excitatory, while fig. 3.10D presents results for the case in which $\mathcal{B}_0$ is inhibitory. The vertical dotted line marks the case of no bias, i.e. $\beta_e = 1$. Otherwise, the meaning of colors and symbols is the same as in all other plots. The overall picture is quite similar to the previous case: the readout scheme B is more effective in detecting the stimulus, provided that the readout is biased. When the stimulated cell is inhibitory, a small weight modification is sufficient to detect the single cell stimulation, while the necessary bias is larger when $\mathcal{B}_0$ is excitatory. The effect size obtained from the readout scheme A is somewhat lower than in the case of biased connections, especially for large bias. The variance of the readout activity σ_A^2 depended on the bias λ rather weakly. To the contrary, if the value of β_e in eq. (3.51) is changed, the variance σ_A^2 changes according to

$$\sigma_A^2(\beta_e) = \sigma_A^2(\beta_e = 0) + (\beta_e^2 - 2\beta_e)\frac{S_{xx}^E \lambda_0}{N_A(1-\lambda_0)\sqrt{\pi}\tau_f}, \tag{3.52}$$

which can be obtained by combining eq. (3.51) with eq. (2.65). Equation (3.52) makes clear that when β_e is large the readout variance grows, thus reducing the effect size.

In summary, adding recurrent excitatory connections to the RN or using an alternative definition of the readout bias, based on the connection strength rather than on the connection probability, does not change the qualitative picture: the single-cell stimulation can be much more easily detected by employing a second readout network with recurrent inhibition rather than with the detection scheme of the previous chapter.

3.5. Summary and discussion

The goal of this chapter was to address some limitations of the readout procedure introduced in chapter 2 by introducing a more biologically realistic detection scheme, which also turned out to be more effective.

One first difference from the detector of the preceding chapter is the replacement of the two-barrier detector with two one-barrier detectors. Although this change may almost sound like a purely lexical substitution, it has two advantages. First, the functional meaning of the detectors with a single barrier is more easily imagined. A reaction triggered by the crossing of an upper barrier might represent an excitatory neuronal population reaching an activity level sufficient to trigger a downstream event. The reaction upon crossing a lower barrier may be seen as an inhibitory population reducing its output so much that its target is activated through disinhibition. A mixed population playing both roles is perhaps possible but less straightforward to interpret at the biological level. Furthermore, the double-barrier detector suffered from a technical weakness (discussed in detail in appendix B) reducing its ability to detect a signal in

one of the two directions.

These considerations suggest that the single-barrier detector is a better representation of how the activity of the readout population can trigger the behavioral effect. Even with this improvement, using the readout scheme A introduced in the previous chapter is equivalent to assuming that the decision about the presence of the stimulus can take place within the stimulated network itself. On the functional level, it seems unlikely that a population in a primary sensory area can trigger a complex motor output as the licking response. Hence, having a second network as the readout circuit represents a significant extension of the model.

The first configuration for the readout network (section 3.2) can be regarded as implementing the single-barrier detector with LIF neurons: the subthreshold dynamics perform a low-pass filtering, only with a filter of a different shape, and there is no difference at all between the firing threshold and the decision threshold (the reset mechanism has no parallel in the detector). As a consequence, it may seem in retrospect not too surprising that the outcome of the two readout schemes is similar. Yet, a large number of threshold units with independent noise can potentially decode a signal better than one single unit, a phenomenon called suprathreshold stochastic resonance (Stocks, 2000). In the case considered here, however, the noise each neuron receives is to a large degree correlated, which makes many readout neurons effectively redundant.

In section 3.3, inhibitory neurons were added to the readout network. Local inhibitory neurons desynchronize the activity of the readout network and reduce cross-correlations in the input (Renart et al., 2010), the main source of noise in the readout activity. If the readout is not biased, or if the bias is exactly the same for both excitatory and inhibitory populations in the readout network, the signal is likewise canceled by inhibition (Hu et al., 2014). Importantly, the signal is strongly influenced by the bias while cross-correlations depend weakly on λ, because the larger contribution to input cross correlations is due to global oscillations and not to shared input (at least for the more relevant case of small λ). These global cross-correlations are removed by inhibition regardless of the bias, thus enabling the readout inhibition to cancel a large portion of noise without eliminating the signal.

As in chapter 2, the bias in the connections is understood as an effect of learning, that is, in biological terms, a result of synaptic plasticity. There is no reason to exclude *a priori* that only one specific type of connection undergoes synaptic plasticity, so that the bias to excitatory and inhibitory populations in the readout network was considered separately and in different combinations. The theoretical and numerical results of section 3.3 show that biasing connections from the stimulated network to the excitatory readout neurons (controlled by the parameter λ_e) is substantially equivalent to biasing connections from the stimulated network to the inhibitory readout neurons (parameter λ_i), provided that the role of the two barriers is interchanged.

Another "anti-symmetry" is found in the role of the two bias parameters in detecting the stimulation of a cell of a particular type: for instance, if the upper-boundary θ_+ detector is used,

stimulating an excitatory cell generates a positive effect size, either if the bias λ_e is increased above λ_0 or if the bias λ_i is reduced below λ_0, and the other way around for the stimulation of an inhibitory cell. In other words, detecting both cell types can be achieved by using only one type of detector (for instance, only θ_+), but it still requires two separate readout networks. Alternatively, one could use a single readout network but then apply both detectors θ_+ and θ_- separately. Biologically, the first solution seems more realistic.

These results were obtained by varying one bias parameter at a time and leaving the other one at its natural value. However, it cannot be ruled out, and it is perhaps even more likely, that synaptic plasticity affects both connections at the same time. If the two bias parameters are changed in parallel, the two possible paths from the barrel cortex network to the readout population $\mathcal{S}^B$, i.e. the direct one and that via the inhibitory population $\mathcal{I}$ work against each other. As a consequence, the signal is suppressed. The converse is true when the two bias parameters are changed in opposite directions: in this case, the effects of the signal traveling through the two paths combine and the effect size is greatly enhanced. One could speculate that the connection bias is the product of a Hebbian-like learning rule, i.e. a rule that tries to maximize correlations between the firing rate of $\mathcal{B}_1$ and of $\mathcal{S}_B$. Potentiating direct connections between these two populations (by increasing λ_e) or weakening the effective feed-forward inhibition (by decreasing λ_i) are two ways of obtaining the same effect. Hence, it seems more likely that such a learning rule would rather change the two bias parameters in opposite directions, than shift them perfectly in parallel, which produces little or no effect.

The aim of section 3.4 was to check the robustness of the results of the previous section. To this end, recurrent excitation was added to the readout network and the number of feed-forward inputs was reduced. The effect of recurrent excitatory connections is, on the one hand, to increase the readout noise by intensifying cross-correlations within the readout network, while on the other hand to amplify the signal. Reducing the number of feed-forward inputs can only reduce the effect size. Nevertheless, the single-cell stimulation remained detectable even if the input coming from the barrel cortex network corresponded to about 25% of the input connections and only 12% of the total excitatory input spikes, if the external shot-noise is taken into account.

Lastly, a different possibility to bias the readout was studied: instead of changing the probability to receive connections from $\mathcal{B}_1$, the average strength of direct connections from $\mathcal{B}_1$ to $\mathcal{S}_B$ was changed while keeping the total mean input to the readout network constant to prevent the mean firing rate of the readout to change significantly. It is conceivable that such a change in the synapses can be realized by a Hebbian learning rule with homeostasis. The results for this last scenario were qualitatively similar to the previous ones.

As long as the readout network is provided with inhibition, the new readout scheme introduced in this chapter lowered the bias required for detection substantially. Put differently, the system needs a much smaller rewiring to learn the task. Notably, *some* learning is still necessary: if the

stimulation were detectable in the model without any bias at all, it would be in disagreement with the experiments, because untrained animals are not able to report single-cell stimulation.

One weakness of the "old" detection procedure (the readout scheme A) that has not been removed by the new readout scheme is that the bias is specific to one particular $\mathcal{B}_1$, and therefore to one $\mathcal{B}_0$, whereas the training is done by microstimulation, which does not target a specific cell (see section 1.2). However, microstimulation pulses were repeated in-between single-cell stimulation trials to maintain the rat in an attentive state (see section 1.2 on p. 14). It is possible that microstimulation can alter the bias dynamically during trials and divert it toward the area around $\mathcal{B}_0$, thus effectively adjusting λ. Strictly speaking, this picture would require a network with spatial structure, which is still not included in the model. Although the dynamics of a network with a space-dependent connectivity profile can be different from those of a random network, it has been observed that global fluctuations are a major source of cross-correlations in cortical networks (Rosenbaum et al., 2017). In such a situation, a readout circuit biased as discussed in this chapter would still be able to detect the stimulation.

The results of this chapter show that stimulating a single-cell in a random recurrent network can induce a transient detectable change in the activity of a second network. Is it realistic to assume that activating a second network is enough to provoke a behavioral response? Anatomical studies show that direct connections from the somatosensory area S1 (to which the barrel cortex belongs) to the primary motor area M1 exist (Feldmeyer et al., 2013). However, these connections target motor areas related to whisker movements (Alloway et al., 2004; Chakrabarti et al., 2008), which are separated from those responsible for tongue control (Miyashita et al., 1994). Therefore, it is possible that triggering the licking response requires a the involvement of (at least) a third processing stage.

3.6. Table of parameters

Table 3.1 reports all parameters used in this chapter with their numerical value.

Symbol	Value	Description
τ_m	20 ms	membrane time constant
τ_{ref}	2 ms	refractory period
v_T	20 mV	threshold voltage
v_R	10 mV	reset voltage
$R_m I_0$	5.2 mV (−18 mV)	constant external input (RN neurons in section 3.2)
C_{ext}	700	number of excitatory external Poisson inputs per neuron
r_{ext}	12 Hz	rate of excitatory external Poisson inputs
N_E	80 000	number of excitatory neurons in the BCN
γ	0.25	ratio of inhibitory to excitatory neurons
N_I	γN_E	number of inhibitory neurons
C_E	4000	number of excitatory inputs per neuron
C_I	γC_E	number of inhibitory inputs per neuron
J	0.1 mV	average synaptic coupling strength
g	7	strength of inhibitory relative to excitatory coupling
D_{min}	0.5 ms	minimum transmission delay
D_{max}	2.0 ms	maximum transmission delay
$\hat{C}$	4000 (1000)	inputs from BCN to RN per neuron (section 3.4)
N_B	10 000	number of excitatory neurons in the RN, i.e. in $\mathcal{S}^B$
$N_{\mathcal{I}}$	γN_B	number of inhibitory neurons in the RN
$C_{\mathcal{E}}$	$C_E - \hat{C}$	number of recurrent excitatory inputs per neuron in the RN
$C_{\mathcal{I}}$	C_I	number of recurrent inhibitory inputs per neuron in the RN
T_s	400 ms	stimulus duration
$R_m \Delta I_0$	23 mV	stimulus intensity
T_w	1200 ms	time window for single-cell detection
τ_f	100 ms	width of time filter for detection
N_A	$\hat{C}$	number of neurons in the readout set $\mathcal{S}^A$
T_{ic}	500 ms	initial simulation time to forget initial conditions
T	3000 ms	simulation time (data acquisition)
Δt	0.1 ms	simulation time step
N_{trials}	900	Number of trials for each network simulation

Table 3.1. – List of parameters used in this chapter with respective numerical value

Chapter 4.

Detecting the Stimulation of a Single Cell in a More Detailed Network Model

After the initial finding that rats can be trained to report the occurrence of single-cell stimulation, Brecht and coworkers investigated in a second series of experiments how the probability of a behavioral response is influenced by the properties of the injected stimulus and of the elicited spike train. In particular, they studied the effect of varying the spike number, the firing rate, and the regularity spike train evoked by the nanostimulation (Doron, 2012; Doron et al., 2014). Their findings can be summarized as follows:[1]

- The total number of elicited spikes was varied by injecting current steps of constant intensity and varying duration. The linear correlation between the average number of evoked spikes and the effect size was slightly negative (fig. 4.1A).

- The firing frequency of the stimulated cell was varied while keeping the total number of elicited spikes approximately constant. To this end, current steps of different length and intensity that are inversely proportional to the duration were used. In other words, the total injected charge was constant. A very weak negative correlation between the firing rate of the stimulated cell and the effect size was observed (fig. 4.1B).

- To evoke irregular spike trains, random permutations of a sequence of current steps of different intensities and durations were used. The irregularity of the elicited spike train was quantified by the coefficient of variation (CV) of the intervals between spikes (see section 1.3). A significant positive correlation between the effect size and the CV of the evoked spike train was found. In other words, irregular spike trains had a better chance to elicit a behavioral response in comparison to a constant current injection. This was the strongest of the three effects (fig. 4.1C).

[1] These results concern the stimulation of excitatory regular-spiking (RS) neurons. Although the same experiments were performed for fast-spiking (FS) inhibitory cells, the paucity of data does not allow for a conclusive statement about the dependence of effect size on the various stimulation parameters.

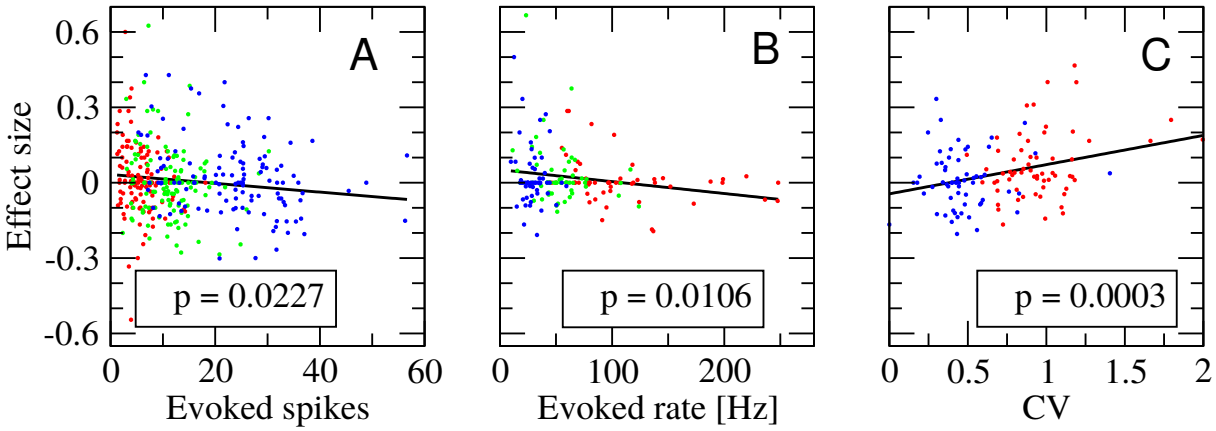

Figure 4.1. – **Summary of experimental results on how properties of the elicited spike train influence the effect size (plots based on dataset from the Brecht lab, key results of Doron, 2012, are reproduced here).** **A**: effect size measured from 119 regular spiking cells (RS) as a function of the average elicited spike count. Colors refer to the stimulation length (red: 100 ms, green: 200 ms, blue: 400 ms). **B**: effect size measured from 55 RS cells as a function of the average evoked firing rate. Colors indicate the stimulation length, the current intensity is inversely proportional to the duration (red: 100 ms, green: 200 ms, blue: 400 ms). **C**: effect size measured from 62 RS cells as a function of the average CV of the evoked spike train. Colors denote the stimulus type (red: random irregular stimulation, see fig. 4.7C, blue: constant current of duration 400 ms).

Taken together, these three experimental findings suggest that the ability to detect the perturbation is rather insensitive to the strength (both in terms of number and frequency of elicited action potentials) of a constant input, while it is more sensitive to fast changes.

The goal of this chapter is to develop a network model able to capture the dependencies summarized above. To this end, the recurrent network model for the surroundings of the stimulated cell will be extended to include several biological details of the barrel cortex. The most notable additions will be: a second class of inhibitory cells (somatostatin-expressing neurons, SOM), short-term plasticity of synaptic connections, spike-frequency adaptation, and some degree of heterogeneity in the cellular parameters of the three different cell classes. One further important difference from the model of the previous chapters is the size, which is here limited to a total of 2600 neurons, the approximate number of neurons within a sphere of about 200 µm around the stimulated cell. Beyond sheer technical convenience (the computational cost to model each neuron and synapse is larger because of the new model features), this choice is also justified by more realism, if a purely random graph is to be considered. As the network connectivity can be considered independent of the distance only on a relatively short range (Avermann et al., 2012; Schnepel et al., 2014), modeling a larger area would require a distance-dependent connectivity profile. Studying the detectability of the single-cell stimulation in a network with spatial structure is a completely new scenario, which will be left for future studies.

Some of the new mechanisms introduced into the model (the short-term depression and the spike-frequency adaptation) provide negative feedback to constant inputs, so that they could be expected to suppress the detectability of a constant stimulus to a much larger degree than in the case of an irregular, changing input. The results presented in the following suggest that these mechanisms alone may not suffice to explain the dependencies seen in the data, if the readout acts as an integrator (with threshold) of the network activity, as in the previous chapters. However, a readout considering *variations* in the network activity yields results that are in much closer agreement to the experimental findings: the effect size barely shows any dependence on the length and intensity of a regular stimulus, and irregular stimuli can be detected more reliably than regular ones. In practice, this new readout considers the difference between the filtered network activity at two different time points. Furthermore, this "differentiation" operation can be approximately implemented using a network of integrate-and-fire neurons by suitably tuning the second readout network architecture introduced in the previous chapter. The fine-tuning of the readout network parameters necessary for its operation as a differentiator readout can be hypothesized as resulting from the training phase.

Interestingly, the bias towards the subset of the network receiving direct input from the stimulated cell, the key factor for detectability in the previous chapters, is not necessary when the stimulated cell is an excitatory regular-spiking (RS) neuron. Stimulating a RS neuron has an *inhibitory* effect on the entire network by engaging the somatostatin-expressing inhibitory neurons, which were missing in the previous models. *In vitro* (Silberberg and Markram, 2007; Kapfer et al., 2007) and even *in vivo* (Kwan and Dan, 2012) experiments suggest that the strong stimulation of a single pyramidal excitatory cell in the barrel cortex has a prevalently inhibitory effect on its surroundings. These studies also demonstrated that this inhibition is due to the action of SOM inhibitory cells and related to the strong facilitation of synapses connecting RS to SOM neurons.

When the stimulated cell is a fast-spiking (FS) inhibitory neuron, the experimental dataset is too small to allow reliable conclusions about the dependence on stimulus parameters, but it does indicate that the detectability is, on average, higher than for RS neurons. In the model studied in this chapter, however, the effect size measured for the stimulation of a FS neuron is generally small and requires a strong readout bias, which is defined analogously to the previous chapters. It is possible that explaining the higher detectability of FS cells requires either a different tuning of the model or a substantially new type of readout or network model.

This chapter is structured as follows. Section 4.1 gives an overview of the model, while its subsections focus on different aspects: section 4.1.1 describes in detail the recurrent network representing the surroundings of the stimulated cell, including the dynamics of synaptic short-term plasticity and of spike-frequency adaptation; section 4.1.2 deals with the properties of the spontaneous network activity and of the simulation procedures; section 4.1.3 studies the firing-

rate response to the stimulation emphasizing the fundamental difference in the response to the stimulation of a RS or a FS neuron; section 4.1.4 describes the three readout schemes used in this chapter (integrator, differentiator, and differentiator network). The main results are presented in section 4.2, which is divided in three subsections, corresponding to the three experiments by Doron et al. (2014): section 4.2.1 investigates the effects of varying the stimulus duration (at constant current); section 4.2.2 studies the dependence on stimulus intensity (with fixed injected charge); and section 4.2.3 compares the detectability of regular and irregular stimuli. Section 4.3 is a summary and discussion of the chapter. The lengthy list of all model parameters is organized in several tables found in section 4.4.

4.1. Model

As in the previous chapter, the model consists of a recurrent network, in which a randomly selected cell is stimulated, and of a readout (fig. 4.2). A first difference from the models considered in chapters 2 and 3 is that the recurrent network represents here only the immediate surroundings of the stimulated cell, in which connection probabilities can be approximated as constant. A reasonable value for the distance over which this assumption may hold is 200 µm (Avermann et al., 2012; Schoonover et al., 2014). Assuming a uniform density of about 79 000 neurons/mm^3 (Meyer et al., 2010), the total number of neurons in the recurrent network can be taken as $N = 2600$. This network size corresponds just to a fraction of one single barrel. Nevertheless, this part of the model will be indicated as "barrel cortex network" (BCN) to be consistent with the previous chapters.

The BCN consists of three populations: one population of excitatory regular spiking cells (RS), one of inhibitory fast-spiking (FS) cells, and one of somatostatin-expressing low-threshold spiking (SOM-LTS) inhibitory cells. These three cell types account for a very large fraction of the neurons in the barrel cortex (about 99% of the neurons in layer IV, Beierlein et al., 2003). In this chapter, the BCN is a more detailed model of the barrel cortex as far as the properties of both single neurons and synaptic connections are concerned. These additional features are: short-term plasticity for synaptic connections, sparse excitatory and dense inhibitory connection probability, spike-frequency adaptation for RS and SOM-LTS neurons, a simple model for the electrical coupling (gap junctions) within the FS and SOM-LTS population, and heterogeneity in several single-neuron parameters. These features are explained in detail in section 4.1.1.

Similarly to chapter 3, the readout receives input from randomly selected excitatory RS cells, and returns the effect size. In this chapter, three possible readout schemes will be examined, as illustrated in fig. 4.2. The first one (fig. 4.2A) is very similar to the readout scheme A of chapter 4, in which a subset of the excitatory neurons of the BCN is selected at random and used as input source to the detector. The main difference from the previous chapter is how spike

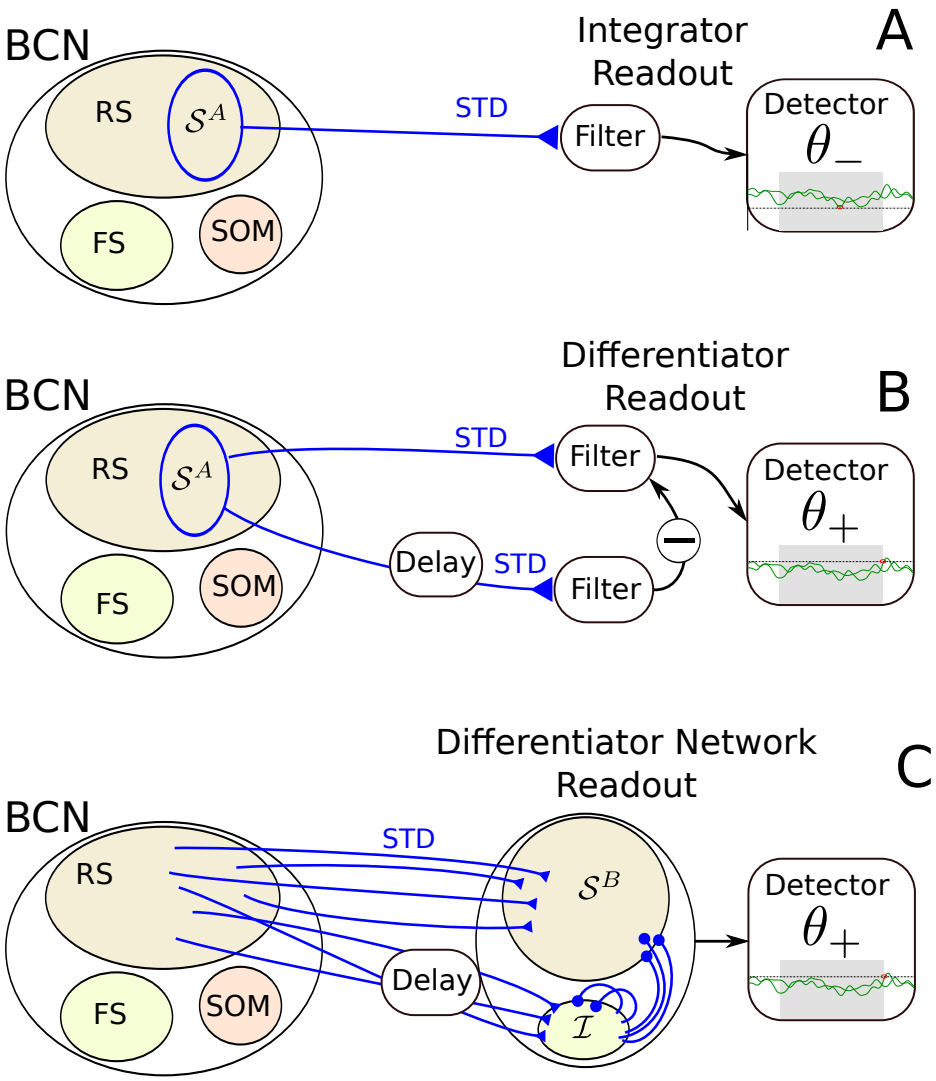

Figure 4.2. – **General model considered in this chapter.** A cell selected at random from the barrel cortex network (BCN) is selected at random and stimulated, as in the previous chapters. In this chapter, the BCN consists of three populations: excitatory regular-spiking neurons (RS), inhibitory fast-spiking neurons (FS), and somatostatin-positive low-threshold spiking neurons (SOM-LTS). Compared to the previous chapters, the BCN includes more biological details (see fig. 4.3). Three readout schemes are considered. **A**: the integrator readout (IR) integrates the activity of a subset of the RS neurons within the BCN. A single-barrier detector extracts the effect size. This scheme is similar to the scheme A of the previous chapter. **B**: the differentiator readout (DR) evaluates the difference between the IR activity at two time points separated by a delay. This filtered running difference at fixed lag is processed by the detector. **C**: the differentiator network readout (DNR) implements the operation of the DR with two populations of LIF neurons. The FS readout population provides delayed recurrent inhibition to itself and feed-forward inhibition to the RS readout population. This readout is similar to the second architecture of the readout scheme B of the previous chapter. All connections depicted in blue are dynamic and show short-term depression (STD).

trains are filtered before entering the detector, which is here a combination of synaptic filtering with short-term depression and of a leaky integration instead of a static filter. This readout scheme will be called integrator readout (IR). The second readout scheme (fig. 4.2B) filters the activity in the same way as the IR, but it subtracts a time-shifted copy of the same activity. In other words, it considers the difference between the filtered activity at different time points, thus acting as a sort of differentiator. For this reason, it will be referred to as differentiator readout (DR). The third readout scheme is the implementation of the DR by means of a simple network of LIF neurons, a differentiator network readout (DNR). It consists of two populations (one excitatory RS population and one inhibitory FS population) and is similar to the second architecture for the readout network considered in the previous chapter. The three readout schemes are described in detail in section 4.1.4.

4.1.1. Barrel cortex network

Figure 4.3 shows a scheme containing all essential features of the BCN, briefly described in the figure caption. The BCN consists of three populations. The largest one counts $N_e = 2000$ excitatory RS neurons, the second one comprises $N_i = 400$ inhibitory FS neurons, and the third one counts $N_s = 200$ SOM-LTS inhibitory neurons, because SOM neurons are, on average, less numerous than FS neurons (Tremblay et al., 2016).

Single-neuron properties and total input to neurons

All neurons are modeled as leaky integrate-and-fire (LIF) neuron models (see section 1.4.1). The kth neuron evolves according to

$$\tau_{m,k}\dot{v}_k(t) = -v_k(t) + R_{m,k}I_{\text{total},k}(t), \tag{4.1}$$

where the membrane time constant $\tau_{m,k}$ is drawn from a lognormal distribution with mean $\tau_{m,e} = \tau_{m,s} = 20\,\text{ms}$ if k is a RS neuron or a SOM-LTS neuron, or with mean $\tau_{m,i} = 10\,\text{ms}$ if k is a FS neuron. The standard deviation of all three distributions was set to 20% of the mean. These values are in rough agreement with experimental values for the rat barrel cortex (Beierlein et al., 2003; Harrison et al., 2015). The membrane resistance is $R_{m,k} = \tau_{m,k}/C_m$, where a capacitance of $C_m = 150\,\text{pF}$ is assumed for all neurons (Harrison et al., 2015). Whenever the voltage reaches the threshold value $v_{T,k}$, the voltage is reset and clamped at $v_R = 10\,\text{mV}$ for the duration of the refractory period $\tau_{\text{ref,k}}$. The threshold voltage is drawn for each neuron from a Gaussian distribution (Harrison et al., 2015) with mean $v_{T,E} = v_{T,I} = 20\,\text{mV}$ if k is an RS or FS neuron (Beierlein et al., 2003; Harrison et al., 2015) and with mean $v_{T,S} = 14\,\text{mV}$ if the kth neuron belongs to the SOM-LTS population, in accordance with the fact that the distance from resting potential to threshold is 5 mV to 7 mV lower in SOM-LTS neurons, compared to

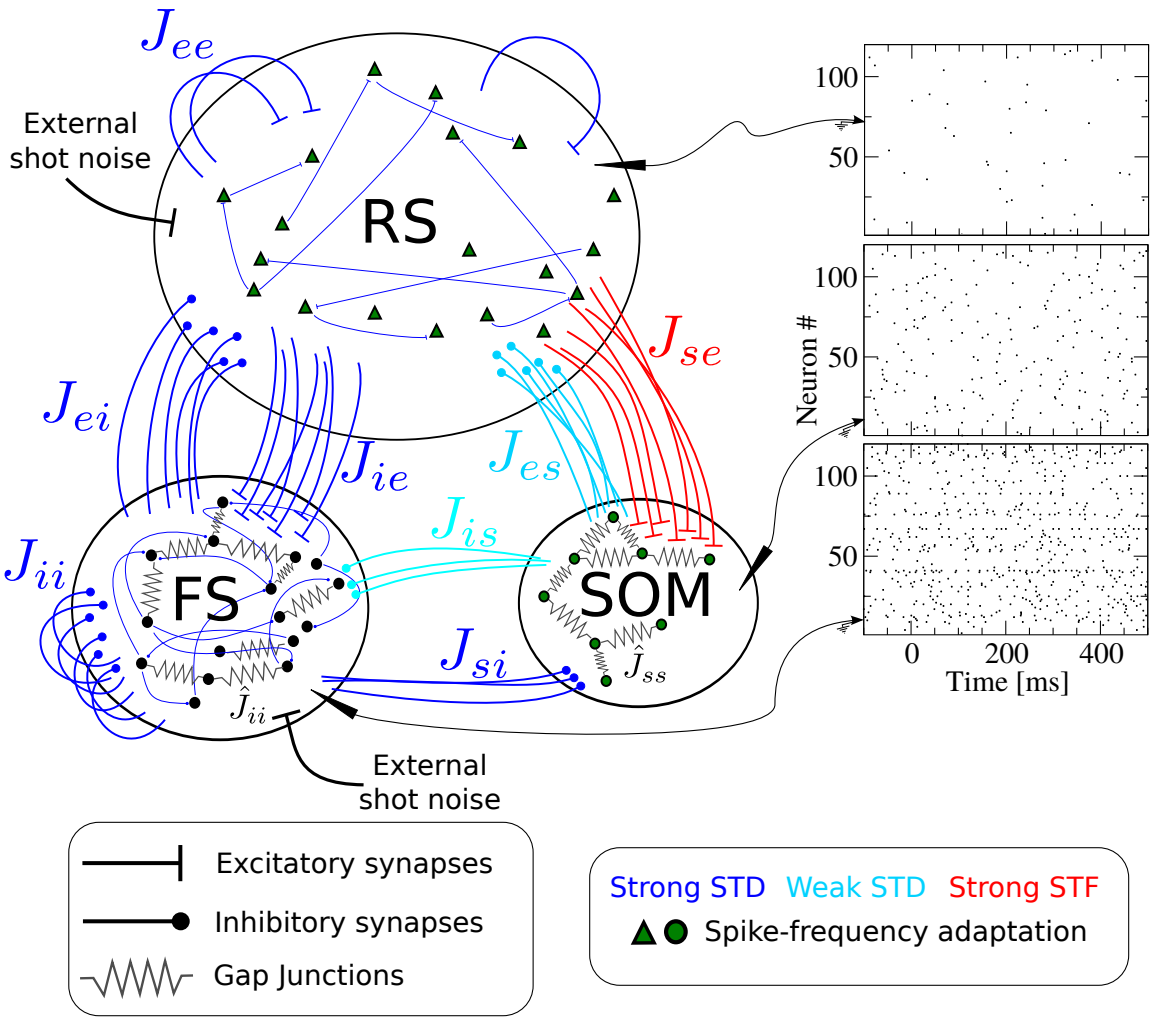

Figure 4.3. – **Recurrent network model representing the surroundings of the stimulated cell.** The network is formed by $N_e = 2000$ excitatory regular spiking (RS) neurons, $N_i = 400$ inhibitory fast spiking (FS) neurons, and $N_s = 200$ inhibitory somatostatin-positive low-threshold spiking (SOM-LTS) neurons. Recurrent connections between RS neurons are sparse (15%), all connections involving FS neurons as well as those between RS and SOM-LTS neurons are dense (40%-50%). FS and SOM-LTS neurons are electrically coupled (only neurons of the same type). Gap junctions are represented by an effective all-to-all spiking coupling (see main text). Connections in blue are strongly depressing, connections in light blue are weakly depressing (see fig. 4.4C,D), and connections in red are strongly facilitating (fig. 4.4E,F). RS and SOM-LTS neurons are endowed with a spike-frequency adaptation current (fig. 4.4A,B). Input from the thalamus and from neighboring cortical regions is represented by Poissonian shot noise. SOM-LTS neurons do not receive external shot noise. The three raster plots show the spontaneous activity of 120 (from top to bottom) RS, SOM, and FS neurons. For all three populations the activity is asynchronous irregular. The spontaneous mean firing rate of excitatory RS neurons is $r_{\mathrm{sp,e}} \approx 0.8\,\mathrm{Hz}$, of SOM-LTS neurons is $r_{\mathrm{sp,s}} \approx 3\,\mathrm{Hz}$, and of FS neurons is $r_{\mathrm{sp,i}} \approx 10\,\mathrm{Hz}$.

that of RS and FS neurons (Beierlein et al., 2003). The standard deviation is 10% of the mean for all three neuron types (Beierlein et al., 2003). The refractory time is $\tau_{\mathrm{ref,k}} = \tau_{\mathrm{ref},0} + \hat{\tau}_{\mathrm{ref,k}}$, where $\tau_{\mathrm{ref},0} = 4\,\mathrm{ms}$ and $\hat{\tau}_{\mathrm{ref,k}}$ is drawn from a lognormal distribution of mean 2 ms and standard deviation 1 ms. The variability in the refractory time was introduced to mimic the experimentally observed variability in the maximum firing rate of neurons (Beierlein et al., 2003).

If the kth neuron belongs to the FS spiking population, its total input current $I_{\mathrm{total},k}$ is just the sum of the external input and of the recurrent input:

$$\underset{k\in\mathrm{FS}}{R_{m,k} I_{\mathrm{total},k}(t)} = R_{m,k}[I_{\mathrm{ext},k}(t) + I_{\mathrm{rec},k}(t)], \tag{4.2}$$

where the first term on the right side of eq. (4.2) represents the input from outside the network, and the second the input from other neurons within the network. When the considered neuron belongs to the RS or to the SOM-LTS population, the total input current has an additional term modeling spike-frequency adaptation.

Spike-frequency adaptation

In the barrel cortex, RS and SOM-LTS neurons show marked spike-frequency adaptation, whereas FS neurons do not (see fig. 4.4B and Gottlieb and Keller, 1997; Beierlein et al., 2003). Therefore, if neuron k belongs either to the RS or to the SOM-LTS population, the total input current includes an additional adaptation term $a_k(t)$:

$$\underset{k\in\mathrm{RS,\,SOM}}{R_{m,k} I_{\mathrm{total},k}(t)} = R_{m,k}[I_{\mathrm{ext},k}(t) + I_{\mathrm{rec},k}(t) - a_k(t)]. \tag{4.3}$$

The adaptation current obeys (Benda and Herz, 2003; Schwalger and Lindner, 2013):

$$\tau_{a,k}\dot{a}_k(t) = -a_k(t) + \tau_{a,k}\Delta a_k x_k(t), \tag{4.4}$$

where $x_k(t) = \sum_j \delta(t - t_{k,j})$ is the spike train emitted by neuron k. In other words, every time the neuron emits a spike, the adaptation current is increased by Δa_k. Between spikes, it relaxes back to zero with a time constant $\tau_{a,k}$. Both Δa_k and $\tau_{a,k}$ are randomly drawn from a lognormal distribution with standard deviation of 20% of the mean. For RS neurons, the means of the two distributions are $\tau_{a,e} = 100\,\mathrm{ms}$ and $\Delta a_e = 0.3\,\mathrm{nA}$, respectively, and for SOM-LTS neurons they are $\tau_{a,s} = 50\,\mathrm{ms}$ and $\Delta a_s = 0.2\,\mathrm{nA}$, respectively. These values are chosen such that the strength of the spike-frequency adaptation is in rough agreement with *in vitro* measurements from the layer IV of the rat barrel cortex, as it can be seen by comparing fig. 4.4A and fig. 4.4B (see the figure caption for more details).

External input to the network

The external input term consists of a constant term and of two sums of excitatory shot-noise processes:

$$R_{m,k}I_{\mathrm{ext},k}(t) = R_{m,k}I_0 + \tau_{m,k}\left[\sum_{j=1}^{C_{\mathrm{ext,th,k}}}\sum_{l} J_{k,j,l}\delta(t-t_{k,j,l}) + \sum_{p=1}^{C_{\mathrm{ext,bc,k}}}\sum_{q} J_{k,p,q}\delta(t-t_{k,p,q})\right]. \quad (4.5)$$

The constant term is $R_{m,k}I_0 = 10\,\mathrm{mV}$ for all neurons. The second term represents the input from the thalamus, and the third mimics the input from the rest of the barrel cortex. Because the thalamus has a higher firing rate, the rate of the Poissonian spike times $t_{k,j,l}$ is $r_{\mathrm{ext,th}} = 10\,\mathrm{Hz}$, while the arrival rate of the spikes $t_{k,p,q}$ is $r_{\mathrm{ext,bc}} = 2\,\mathrm{Hz}$. The number of input spike trains depends on the cell type. If k belongs to the SOM-LTS population, then the number of inputs is zero $C_{\mathrm{ext,th,k}} = C_{\mathrm{ext,bc,k}} = 0$, i.e. SOM-LTS do not receive external shot-noise input at all, which is consistent with the experimental observations that SOM cells receive very little input from the thalamus and from distant brain regions (Gibson et al., 1999; Beierlein et al., 2003). If k is a RS or a FS neuron, then $C_{\mathrm{ext,th,k}} = 500$, because the input from the thalamus targets both RS and FS cells (Gibson et al., 1999; Beierlein et al., 2003). Finally, the number of inputs from the cortical surroundings is $C_{\mathrm{ext,bc,e}} = 2000$ when k is a RS neuron, and $C_{\mathrm{ext,bc,i}} = 1000$ when k is a FS neuron, because dendrites of FS neurons tend to be more localized and to receive more input from local RS neurons and less from distant ones. The amplitude of each input spike is drawn independently from an exponential distribution with mean $J_{\mathrm{ext,e}} = 0.1\,\mathrm{mV}$ when k is a RS neuron, and from an exponential distribution with mean $J_{\mathrm{ext,i}} = 0.2\,\mathrm{mV}$ when k is a FS neuron, consistent with the fact that both thalamic and cortical excitatory postsynaptic potential (EPSP) amplitudes are larger in FS cells than in RS cells (Beierlein et al., 2003).

Recurrent input to RS neurons

The recurrent input term $I_{\mathrm{rec},k}(t)$ depends on the identity of the neuron. For excitatory RS neurons, it reads

$$\begin{aligned} R_{m,k}I_{\mathrm{rec,k}}(t) = \tau_{m,k}\Bigg[&\sum_{i\in\mathcal{P}_e(k)} J_{ki}(t)x_i(t-D_{ki}) \\ &- \sum_{j\in\mathcal{P}_i(k)} J_{kj}(t)x_j(t-D_{kj}) - \sum_{\ell\in\mathcal{P}_s(k)} J_{k\ell}(t)x_\ell(t-D_{k\ell})\Bigg], \end{aligned} \quad (4.6)$$

where $x_i(t-D_{ki})$ is the spike train fired by neuron i, D_{ki} is the transmission delay from neuron i to neuron k, J_{ki} is the synaptic strength from neuron i to neuron k, which is a function of time (see below). The term D_{ki} represents the total delay resulting from the axonal propagation,

the neurotransmitter diffusion, and the dendritic propagation. Neuron k receives input from three sets of neurons: $\mathcal{P}_e(k)$, formed by $C_{ee} = 300$ randomly selected RS neurons, $\mathcal{P}_i(k)$ are $C_{ei} = 200$ randomly selected FS neurons, and $\mathcal{P}_s(k)$ are $C_{es} = 100$ randomly selected SOM-LTS neurons. In other words, the probability of a synapse incoming from another RS neuron is $C_{ee}/N_e = 15\%$, while the probability of receiving input from a randomly chosen FS or SOM neuron is $C_{ei}/N_i = C_{es}/N_s = 50\%$, which is in line with the experimental observations that the probability of a connection between RS cells is in the range 5% to 25% (Beierlein et al., 2003; Lefort et al., 2009; Avermann et al., 2012) whereas the probability of a connection from inhibitory neurons to RS neurons is much higher (Beierlein et al., 2003; Silberberg and Markram, 2007; Packer and Yuste, 2011; Avermann et al., 2012; Koelbl et al., 2015). Transmission delays are drawn uniformly in the range 0.5 ms to 1.0 ms (Koelbl et al., 2015). All synaptic weights in eq. (4.6) are not static, but obey a differential equation describing short-term depression (STD):

$$J_{ki}(t) = J_{ki} R_{ki}(t^-), \tag{4.7}$$

where J_{ki} is the value of the maximum strength of the connection, achieved when the presynaptic neuron i has not been firing for a long time, and $R_{ki}(t)$ represents the fraction of available synaptic resources (Tsodyks and Markram, 1997). The notation t^- indicates that the function is evaluated just before a spike is considered (see below). The model for the STD and its parameters are described in detail below. The maximum synaptic couplings J_{ki} are drawn independently for each connection from an exponential distribution. The mean of the distribution for RS-to-RS coupling is $J_{ee} = 0.1\,\mathrm{mV}$, for FS-to-RS coupling is $J_{ei} = 0.5\,\mathrm{mV}$, and for for SOM-to-RS coupling is $J_{es} = 0.25\,\mathrm{mV}$.

Recurrent input to FS neurons

The recurrent input to a FS neuron is:

$$\begin{aligned} \underset{k\in\mathrm{FS}}{R_{m,k} I_{\mathrm{rec,k}}(t)} = \tau_{m,k} \Bigg[& \sum_{i\in\mathcal{Q}_e(k)} J_{ki}(t) x_i(t - D_{ki}) - \sum_{j\in\mathcal{Q}_i(k)} J_{kj}(t) x_j(t - D_{kj}) \\ & - \sum_{p\in\mathcal{Q}_s(k)} J_{kp}(t) x_p(t - D_{kp}) + \sum_{\ell\in FS} \hat{J}_{k\ell} x_\ell(t - D_{k\ell}) \Bigg], \end{aligned} \tag{4.8}$$

where the first two terms represent the synaptic input from RS and FS neurons, respectively. The size of the excitatory RS presynaptic population $\mathcal{Q}_e(k)$ is $C_{ie} = 800$, which corresponds to a connection probability of is $C_{ie}/N_e = 40\%$, the size of the inhibitory FS presynaptic population $\mathcal{Q}_i(k)$ is $C_{ii} = 200$, i.e. the connection probability from FS to FS is $C_{ii}/N_i = 50\%$, and the number of inputs from the SOM-LTS population ($\mathcal{Q}_s(k)$) is is $C_{is} = 50$ (connection probability is 25%). The RS-to-FS, FS-to-FS, and SOM-to-FS synaptic connections obey eq. (4.7), and

their peak value is drawn from an exponential distribution of mean $J_{ie} = 0.2\,\mathrm{mV}$, $J_{ii} = 1.0\,\mathrm{mV}$, and $J_{is} = 0.1\,\mathrm{mV}$ respectively. Transmission delays are the same as for RS-to-RS connections. These values were chosen to model the strong and dense connections that FS neurons receive both from RS and from FS neurons (Beierlein et al., 2003; Pfeffer et al., 2013). Synapses from SOM neurons to FS neurons are weaker in comparison (Pfeffer et al., 2013).

The last term in eq. (4.8) is an effective model for the electrical coupling among FS cells, i.e. mediated by gap junctions.

Effective model for gap junctions

Several experimental studies have shown that both FS and SOM neurons in the rat visual and somatosensory cortex are densely connected with gap junctions, i.e. channels that directly connect the intracellular space of two neurons (Galarreta and Hestrin, 1999; Gibson et al., 1999; Beierlein et al., 2000; Amitai et al., 2002). In a simplified view, these channels act as a passive conductance coupling the membrane voltage of two neurons. An established way of including the electrical coupling of gap junctions into integrate-and-fire networks would be to model both the sub-threshold and the spiking part of the coupling, which means that the effect of the activity of neuron ℓ on neuron k mediated by the gap junction would be (Lewis and Rinzel, 2003; Ostojic et al., 2009b):

$$R_{m,k} I_{GJ,k\ell} = \gamma_{k\ell}(v_\ell - v_k) + \tau_{m,k}\hat{J}_{k\ell} x_\ell(t - D_{k\ell}), \tag{4.9}$$

where $\gamma_{k\ell}$ would be proportional to the Ohmic conductance between the two neurons, and $\hat{J}_{k\ell}$ models the effect of spikes fired by neuron ℓ, which has to be added *ad hoc*, because LIF neurons do not explicitly generate action potentials. The delay term $D_{k\ell}$ is justified by the fact that gap junctions are often formed between dendrites (Tamás et al., 2000). Although the transmission through the gap junction itself is very fast, the time necessary for a spike to propagate along the dendrite of the firing cell to the gap junction and from the gap junction to the soma of the receiving neuron can be as large as 0.5 ms (Tamás et al., 2000). Here, the delays in the gap-junction transmission were drawn uniformly in the range 0.1 ms to 0.5 ms. Furthermore, the sub-threshold coupling was be completely neglected, i.e. in the present network model $\gamma_{j\ell} = 0$ is set for all neuron pairs. Holzbecher and Kempter (2018) demonstrated that the sub-threshold coupling has a very weak effect on the firing rate, synchrony, and oscillation frequency of a network of LIF neurons, whereas the spike-related coupling (here indicated by $\hat{J}_{k\ell}$) has a much larger impact. The size of gap-junction potentials measured in FS neurons of the rat somatosensory cortex was found to be quite variable and, on average, about half as large as excitatory post-synaptic potentials generated by RS neurons (Tamás et al., 2000; Sun et al., 2006). Therefore, each gap-junction coupling term $\hat{J}_{k\ell}$ was drawn from an exponential distribution of mean $\hat{J}_{ii} = J_{ie}/2 = 0.05\,\mathrm{mV}$. The probability of two neighboring inhibitory

neurons of the same type (FS with FS and SOM with SOM) being electrically coupled is high (60% to 80% Gibson et al., 1999; Amitai et al., 2002). For simplicity, the gap-junction coupling was assumed here to be all-to-all (excluding self-coupling).

Recurrent input to SOM-LTS neurons

Finally, the recurrent input in the case that neuron k is a SOM-LTS neuron is

$$\begin{aligned} R_{m,k} \underset{k\in\mathrm{SOM}}{I_{\mathrm{rec,k}}}(t) = \tau_{m,k} \Bigg[& \sum_{i\in\mathcal{L}_e(k)} J_{ki}(t) x_i(t - D_{ki}) \\ & - \sum_{j\in\mathcal{L}_i(k)} J_{kj}(t) x_j(t - D_{kj}) + \sum_{\ell\in SOM} \hat{J}_{k\ell} x_\ell(t - D_{k\ell}) \Bigg], \end{aligned} \tag{4.10}$$

where the three terms have the same meaning as in the previous case: the first is the input from excitatory RS neurons, the second is the input from inhibitory FS neurons, and the third is the coupling due to gap junctions. The effective representation of gap-junctions is implemented here exactly in the same way as in eq. (4.8): it is a global spiking coupling with static amplitudes drawn from an exponential distribution with mean $\hat{J}_{ss} = \hat{J}_{ii} = 0.05\,\mathrm{mV}$ and with short delays distributed uniformly in the interval 0.1 ms to 0.5 ms. The first term in eq. (4.10) is the input from $\mathcal{L}_e(k)$, a randomly chosen set of $C_{se} = 1000$ RS neurons (which corresponds to a connection probability of $C_{se}/N_e = 50\%$). These connections are the only ones that exhibit short-term *facilitation* instead of depression, and for which a stochastic term modeling *transmission failures* was implemented. As in all other cases, the static prefactors J_{ki} modulating the baseline amplitude of each synapse are drawn independently from an exponential distribution and have mean $J_{se} = 0.1\,\mathrm{mV}$. The second term in eq. (4.10) models the input from $C_{si} = 100$ randomly selected FS neurons, indicated by $\mathcal{L}_i(k)$, which corresponds to a connection probability of $C_{si}/N_i = 25\%$. These connections have an average strength of $J_{si} = 0.25\,\mathrm{mV}$, they show short-term depression and obey eq. (4.7). Inhibitory chemical coupling between SOM neurons is not included because it has been found to be very weak (Gibson et al., 1999; Beierlein et al., 2003).

Model of short-term depression

As mentioned above, the strength of all synapses in the BCN is time-dependent. Except for those connecting RS to SOM neurons, all chemical synapses in the BCN model display short-term depression (STD). Each synaptic weight obeying STD dynamics $J_{kj}(t)$ has a time dependence described by

$$J_{ki}(t) = J_{ki} R_{ki}(t^-), \tag{4.11}$$

where the variable $R_{ki}(t)$ describes the fraction of available synaptic resources, i.e. the vesicles containing the neurotransmitter. According to a model introduced by Tsodyks and Markram (1997), the evolution of the variable $R_{ki}(t)$ can be described by the following equation:

$$\dot{R}_{ki}(t) = \frac{1 - R_{ki}(t)}{\tau_D} - U_{se} R_{ki}(t^-) \sum_j \delta(t - \hat{t}_{i,j}), \tag{4.12}$$

where $\hat{t}_{i,j}$ are the times at which the spikes of neuron i arrive at the synapse, and t^- indicates that the function is evaluated at $t - \varepsilon$ (where $\varepsilon > 0$ is an infinitesimally small positive number), i.e. just before a spike. The parameter U_{se} determines the fraction of synaptic used by each release, while τ_D is the time scale with which synaptic uptake mechanisms restore the neurotransmitter stocks. Note that the time evolution of $R_{ki}(t)$ depends on the spike times of the presynaptic neuron i only. Hence, as long τ_D and U_{se} do not depend on k, the time course of each variable $R_{ki}(t)$ is a time-shifted copy of a single master variable $R_i(t)$

$$R_{ki}(t) = R_i(t - D_{ki}), \tag{4.13}$$

where $R_i(t)$ obeys the same equation as $R_{ki}(t)$, except that the arrival times $\hat{t}_{i,j}$ in eq. (4.12) are replaced by $t_{i,j}$, the spike times of neuron i. Here, it is assumed that the parameters τ_D and U_{se} can only depend on the type of the source and target neuron, but not on the identity of the particular neuron within a population. Therefore, eq. (4.13) holds. By exploiting eq. (4.13) the number of actual dynamic variables needed to simulate the network is reduced from one variable per synapse to one variable per neuron, i.e. a decrease by a factor ~ 1000 in the number of necessary variables, a tremendous computational advantage.

The parameter values chosen to model a strong depression are $\tau_D = \tau_{D,s} = 150\,\text{ms}$ and $U_{se} = U_{se,s} = 0.2$ and apply to all synapses marked in blue in fig. 4.3, which are all chemical synapses except for those connecting RS to SOM neurons, and all outgoing synapses of the SOM-LTS neurons. Figure 4.4C shows with black circles the amplitude of a train of post-synaptic potentials (normalized to the first) resulting from a pre-synaptic regular spike train with frequency 40 Hz. It can be seen that the amplitude decreases and the eighth PSP is about one half of the maximal amplitude, which is in rough agreement with fig. 4.4D, which reports *in vitro* measurements of the relative amplitudes of a PSP train transmitted by a RS-to-FS synapse in the layer IV of the barrel cortex in response to 40 Hz pre-synaptic stimulation (adapted from Beierlein et al., 2003). Experimental measurements suggest that most chemical synapses in the barrel cortex are depressing (Beierlein et al., 2003; Helmstaedter et al., 2008; Lefort and Petersen, 2017). However, inhibitory synapses originating from SOM-LTS neurons and terminating onto RS neurons show only weak depression or even slight facilitation. Here, these connections are modeled as depressing (fig. 4.3, light blue), but only weakly: choosing $\tau_D = \tau_{D,w} = 50\,\text{ms}$ and

$U_{se,w} = 0.05$ causes only a moderate reduction of the PSP amplitude to $\approx 95\%$ after eight PSP at 40 Hz (fig. 4.4C, red squares). For simplicity, also SOM-to-FS connections were given the same STD parameters.

Short-term facilitation and transmission failures

Several experimental studies have found that excitatory synapses connecting RS neurons to SOM-LTS neurons (marked in red in fig. 4.3) are strongly facilitating (Beierlein et al., 2003; Silberberg and Markram, 2007; Kapfer et al., 2007). Biologically, synaptic facilitation is probably related to a spike-triggered calcium inflow into the synapse (Cowan et al., 2003). The model by Tsodyks and Markram (1997) was extended to model facilitating synapses by rendering the parameter U_{se} a dynamical variable, $u(t)$ (Markram et al., 1998; Tsodyks et al., 1998). In this model, the amplitude of the PSPs is proportional to the product $R(t)u(t)$. Considering again the connection from neuron i to neuron k, the model for facilitation used here is (note that the conventions have been slightly changed with respect to those of Tsodyks et al., 1998):

$$J_{ki}(t) = J_{ki} R_i(t^- - D_{ki}) \frac{u_i(t^+ - D_{ki})}{U_b}, \tag{4.14}$$

where t^+ indicates that the function is evaluated at $t + \varepsilon$, that is, the value of $u_i(t)$ just after a spike is considered. The two variables $R_i(t)$ and $u_i(t)$ obey the following equations:

$$\dot{u}_i(t) = \frac{U_b - u_i(t)}{\tau_F} + U \cdot \Big(1 - u_i(t^-)\Big) \sum_j \delta(t - t_{i,j}) \tag{4.15}$$

$$\dot{R}_i(t) = \frac{1 - R_i(t)}{\tau_D} - u_i(t^-) R_i(t^-) \sum_j \delta(t - t_{i,j}), \tag{4.16}$$

where $t_{i,j}$ are, as usual, the spike times of neuron i. The first term in eq. (4.15) describes the relaxation to the baseline level U_b (note that Tsodyks et al., 1998, set $U_b = 0$), while the second term causes an upward jump upon each pre-synaptic spike. The first jump has size U, while later jumps are reduced by the term $1 - u_i(t)$, which ensures that u_i does not exceed unity. The time evolution of $R_i(t)$ is analogous to that of a purely depressing synapse, with the only change that the synaptic use parameter is here time-dependent. The combination of the parameters U, τ_F, τ_D determines whether the overall effect for a given pre-synaptic firing rate is facilitating, depressing, or both. Here, the facilitation and depression time scales were set at $\tau_F = 300$ ms and $\tau_D = \tau_{D,f} = 100$ ms, respectively. The baseline value for $u_i(t)$ was set at $U_b = 0.01$ and the increment at $U = 0.03$. With this choice of parameters, the behavior of the synapse described by eq. (4.14) for a pre-synaptic stimulation of 40 Hz is purely facilitating, as depicted in fig. 4.4E (black circles).

Another property that distinguishes RS-to-SOM synapses from all other synapses considered

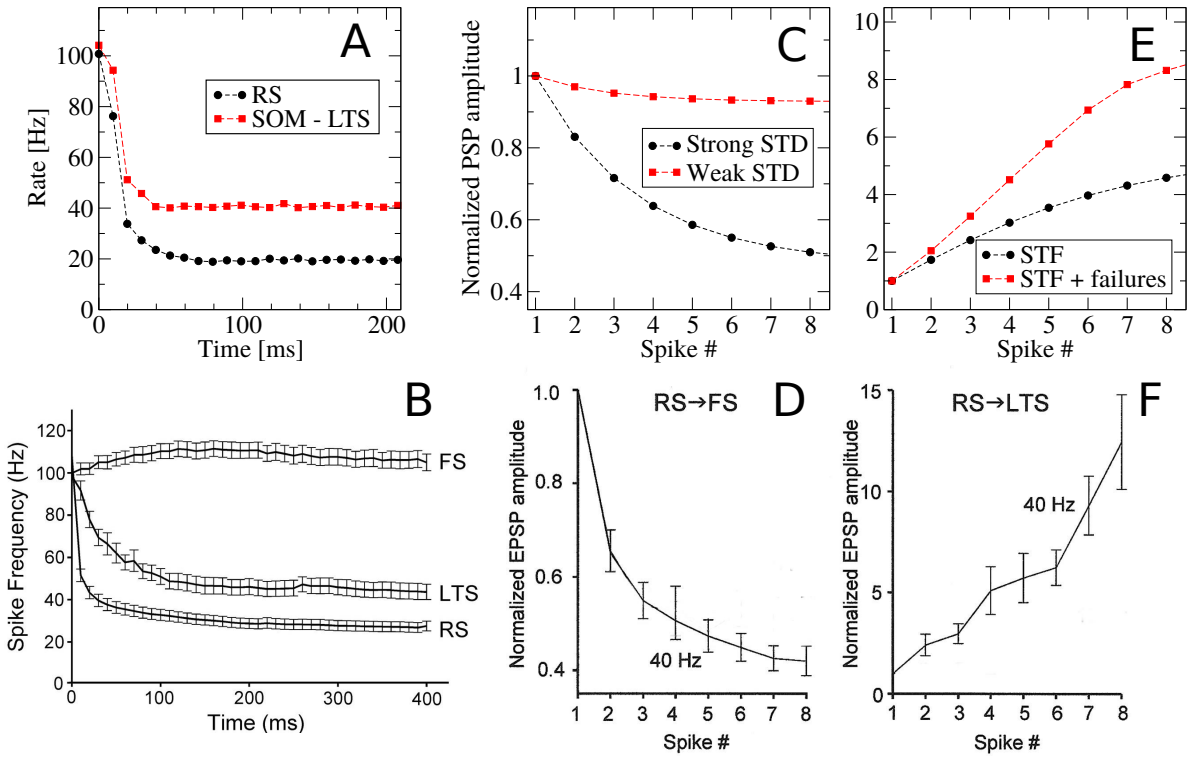

Figure 4.4. – **Spike-frequency adaptation and short-term plasticity qualitatively agree with *in vitro* measurements in the layer IV of the rat barrel cortex**. **A**: Effect of spike-frequency adaptation on the response of a population of leaky integrate-and-fire neurons to a current step. Cellular parameters correspond to those of the RS (black circles) and of the SOM-LTS (red squares) neurons (see table 4.2). The current pulse height is adjusted to elicit an initial firing rate of $\approx 100\,\mathrm{Hz}$. Neurons receive no external input except for a weak background Gaussian white noise representing channel noise ($D = 0.1\,\mathrm{mV^2ms}$). These curves reproduce qualitatively the average firing-rate response measured by Beierlein et al. (2003) in layer IV barrel cortex slices, shown in **B**. As it can be seen, RS display the strongest effect, while FS do not show spike-frequency adaptation. **C**: Effect of short-term depression (STD) on the amplitude of the post-synaptic potentials elicited by a 40 Hz regular spike train. The strong STD applies to all synapses except for connections both from and to the SOM cells (blue connections in fig. 4.3). The weak STD applies to connections originating from SOM cells (light blue connections in fig. 4.3). Amplitudes are normalized to the first peak in the PSP train. **D**: Effect of STD on RS-to-FS synapses in the layer IV of the barrel cortex, measured under the same conditions as in **C**. FS-to-RS synapses also display similarly strong STD (not shown, see Beierlein et al., 2003), and RS-to-RS connections are also mostly depressing (Lefort and Petersen, 2017). **E**: Effect of short-term facilitation (STF) on the amplitude of the post-synaptic potentials elicited by a spike train of frequency 40 Hz. Black circles show only the effect of the facilitation dynamics for the amplitude, i.e. according to eq. (4.14). Red squares include the average effect of transmission failures, i.e. according to eq. (4.20), where the average effect of $S(p_f)$ is considered. **F**: effect of STF on PSP amplitudes of RS-to-SOM-LTS connections measured in layer IV of the barrel cortex. **B, D, and F** were adapted from Beierlein et al. (2003), ©(2003) the American Physiological Society.

here is the much higher occurrence of synaptic transmission failures. Failure rates for other synapses in the barrel cortex have been found to be generally low and to be barely affected by repeated stimulation (average failure rate is $\approx 10\%$ for RS-to-RS synapses, $\approx 5\%$ for synapses to and from FS neurons Beierlein et al., 2003). However, the transmission failure rate of RS-to-SOM synapses is quite large at rest ($\gtrsim 50\%$), and decreases to $\approx 10\%$ upon repeated stimulation at 40 Hz (Beierlein et al., 2003).

Here, transmission failures will be modeled only for RS-to-SOM synapses by introducing a new stochastic binary variable $S(p_f)$:

$$S(p_f) = \begin{cases} 1 & \text{with probability } 1 - p_f \\ 0 & \text{with probability } p_f \end{cases}, \tag{4.17}$$

where p_f is the failure rate. The failure rate at rest in the model is set at $p_{f,\text{rest}} = 0.5$. Every time the presynaptic neuron spikes, the failure rate for the synapse is decreased by $\Delta p_f = 0.1$. The failure rate relaxes back to the baseline value with the time constant $\tau_f = 250\,\text{ms}$. The time evolution of $p_f(t)$ is described by the equation

$$\dot{p}_f(t) = \frac{p_{f,\text{rest}} - p_f(t)}{\tau_f} - G(p_f, \Delta p_f, p_{\min}) \sum_j \delta(t - t_{i,j}), \tag{4.18}$$

where the piecewise linear function $G(p_f, \Delta p_f, p_{\min})$ ensures that the failure rate $p_f(t)$ cannot decrease below the minimum value $p_{\min} = 0.1$. It is defined as

$$G(p_f, \Delta p_f, p_{\min}) = \begin{cases} 0 & \text{if } p_f \leq p_{\min} \\ p_f - p_{\min} & \text{if } p_{\min} < p_f < p_{\min} + \Delta p_f \\ \Delta p_f & \text{if } p_{\min} + \Delta p_f \leq p_f \end{cases}. \tag{4.19}$$

In the end, the synaptic weight from the RS neuron i to the SOM-LTS neuron k obeys the following equation:

$$J_{ki}(t) = J_{ki} R_i(t^- - D_{ki}) \frac{u_i(t^+ - D_{ki})}{U_b} S\left(p_{f,i}(t^- - D_{ki})\right). \tag{4.20}$$

When the average effect of synaptic failures is considered, the increase in the average synaptic amplitude in response to a 40 Hz presynaptic stimulation is greatly increased, as plotted in fig. 4.4E (red squares). The eightfold increase in the PSP amplitude after eight spikes is still somewhat below the experimentally measured twelvefold amplification observed in fig. 4.4F, but it still is a reasonable agreement, considering the large error bars, the variability across layers (for instance, the facilitation measured by Kapfer et al., 2007, in superficial layers is strong but

not as large as in fig. 4.4F), and that the synaptic responses *in vivo* may be different from those measured in *in vitro* (Borst, 2010).

4.1.2. Averaging ensembles, spontaneous activity, and single-cell stimulation

In every trial, the network is initialized at random and simulated for $T_{\text{idle}} = 1200\,\text{ms}$, to let the system approach the stationary state. The spontaneous firing regime of all three populations of the network is asynchronous and irregular (fig. 4.3). The mean spontaneous firing rate of RS, FS, and SOM-LTS neurons is $r_{\text{sp,e}} \approx 0.8\,\text{Hz}$, $r_{\text{sp,i}} \approx 10\,\text{Hz}$, and $r_{\text{sp,s}} \approx 3\,\text{Hz}$, respectively. These values are consistent with experimental observations that the firing rates of pyramidal neurons are low, that inhibitory FS neurons have typically much higher firing rates, and that SOM-LTS neurons have intermediate firing rates (Middleton et al., 2012; Gentet et al., 2012).

As in the previous chapters, a neuron (labeled as $\mathcal{B}_0$) is randomly selected as site of the nanostimulation, which is switched on at $t = 0$ and modeled as additional external current. In this chapter, different stimulus durations and intensities are used. The maximum stimulation current used for RS neurons is $\Delta I_{\text{max},e} = 5\,\text{nA}$, and $I_{\text{max},i} = 2.5\,\text{nA}$ for FS neurons. In the experimental dataset, no data for SOM-LTS neurons are present. Therefore, only the stimulation of RS and FS neurons was considered. After the stimulus is switched off, the network is simulated until the time reaches $t = T_{end} = 1200\,\text{ms}$.

Following Doron (2012); Doron et al. (2014), step currents of different lengths (ranging from 100 ms to 400 ms) and intensities were used to study the dependence of the detectability on the spike count and firing rate of the evoked spike train. Random permutations of a sequence of different steps are used to generate irregular spike trains. Details on the stimuli are provided in section 4.2. In chapters 2 and 3, false positives were extracted from the network activity preceding the stimulus onset. Here, false positive rates were computed by letting the detector act onto dedicated catch trials, i.e. trials in which no stimulus was present. This procedure is closer to the experimental one, and it is safer with respect to a possible residual non-stationarity due to the initial conditions. Two equally sized sets of catch trials were simulated to estimate the size of random fluctuations in the detection rates.

Realizations of the external shot noise and random initial conditions were drawn anew for every single simulation. The same realization of the fixed disorder (randomized cellular parameters, network connectivity including weights and delays) was used once for each stimulus type, including the catch trials. This procedure corresponds to an experiment in which each cell is used for a single presentation of each stimulus type. Results are based on $N_{\text{trials}} = 10000$ network realizations.

4.1.3. Firing-rate response

As mentioned above, the readout receives input from the RS population. Before describing the three readout mechanisms in detail, it is convenient to examine the effect of the single-cell stimulation on the trial-averaged firing rate of the RS neurons. In the BCN model considered in the last two chapters, stimulating an excitatory or an inhibitory cell had an effect that was different in magnitude and opposite in sign, but similar in nature: $\mathcal{B}_0$ would raise or decrease the firing rate of its direct targets (the $\mathcal{B}_1$ population), while the inhibitory feedback would try to compensate by pushing the firing rate of all other neurons (the $\mathcal{B}_2$ population) in the opposite direction. Here, the stimulation of a RS or of a FS neuron activates quite different network paths, which are highlighted in fig. 4.5.

When $\mathcal{B}_0$ is chosen from the RS population, the RS neurons receiving direct input from it are a quite small fraction of the total (15%), while first neighbors of $\mathcal{B}_0$ within the FS and SOM population are 50% of the respective population (represented by the shaded areas in fig. 4.5A). The firing-rate response of the RS neurons to a nanostimulation step is plotted in fig. 4.5a2 (averaged over 10000 trials, spikes are filtered with an exponential filter with decay constant $\tau_f = 15\,\text{ms}$). Just after the stimulation onset, a small peak is observed, which is due to the spikes generated by $\mathcal{B}_0$ itself, as it can be seen by leaving out only $\mathcal{B}_0$ from the average population firing rate (fig. 4.5a2). The peak disappears after about 10 ms because of the combined effect of the spike-frequency adaptation and of the buildup of inhibitory input from the SOM population. The firing rate of the SOM population is shown in fig. 4.5a1. It can be seen how the facilitating synapses lead to a rapid increase in the firing rate of the SOM cells, which peaks about 30 ms after the stimulus onset. Then, it relaxes to a plateau after a mild sag around $t = 100\,\text{ms}$ (a damped oscillation due to spike-frequency adaptation). When the stimulus is switched off, the firing rate of the SOM neurons returns to the baseline. Note how the time course of the RS neurons' firing rate closely follows that of the SOM neurons' activity with opposite sign, except for the overshoot of the RS population's activity, caused by the the spike-frequency adaptation. The response of the RS population is consistent with results of *in vitro* experiments showing that inducing tonic firing in a single excitatory pyramidal cell in the barrel cortex (a RS neuron) has a predominantly inhibitory effect of the surrounding pyramidal cells, and that the effect is mediated by one subclass of SOM cells (Silberberg and Markram, 2007).

When $\mathcal{B}_0$ is an inhibitory FS cell, neurons receiving direct input from it are, on average, 50% of the RS, 50% of the FS, and 25% of the SOM population (fig. 4.5B). The response of all direct targets of $\mathcal{B}_0$ within the RS population ($\mathcal{B}_1^{RS}$) is, intuitively, a reduction in the firing rate, which displays a minimum at $t \approx 40\,\text{ms}$ and a recovery due both to the STD of the inhibitory synapses from FS neurons and to the spike-frequency adaptation (fig. 4.5b1, orange line). However, half of the FS population also receives direct inhibition ($\mathcal{B}_1^{FS}$), which decreases their firing rate (fig. 4.5b2, orange line). As a consequence of this disinhibition, other FS neurons, which do not

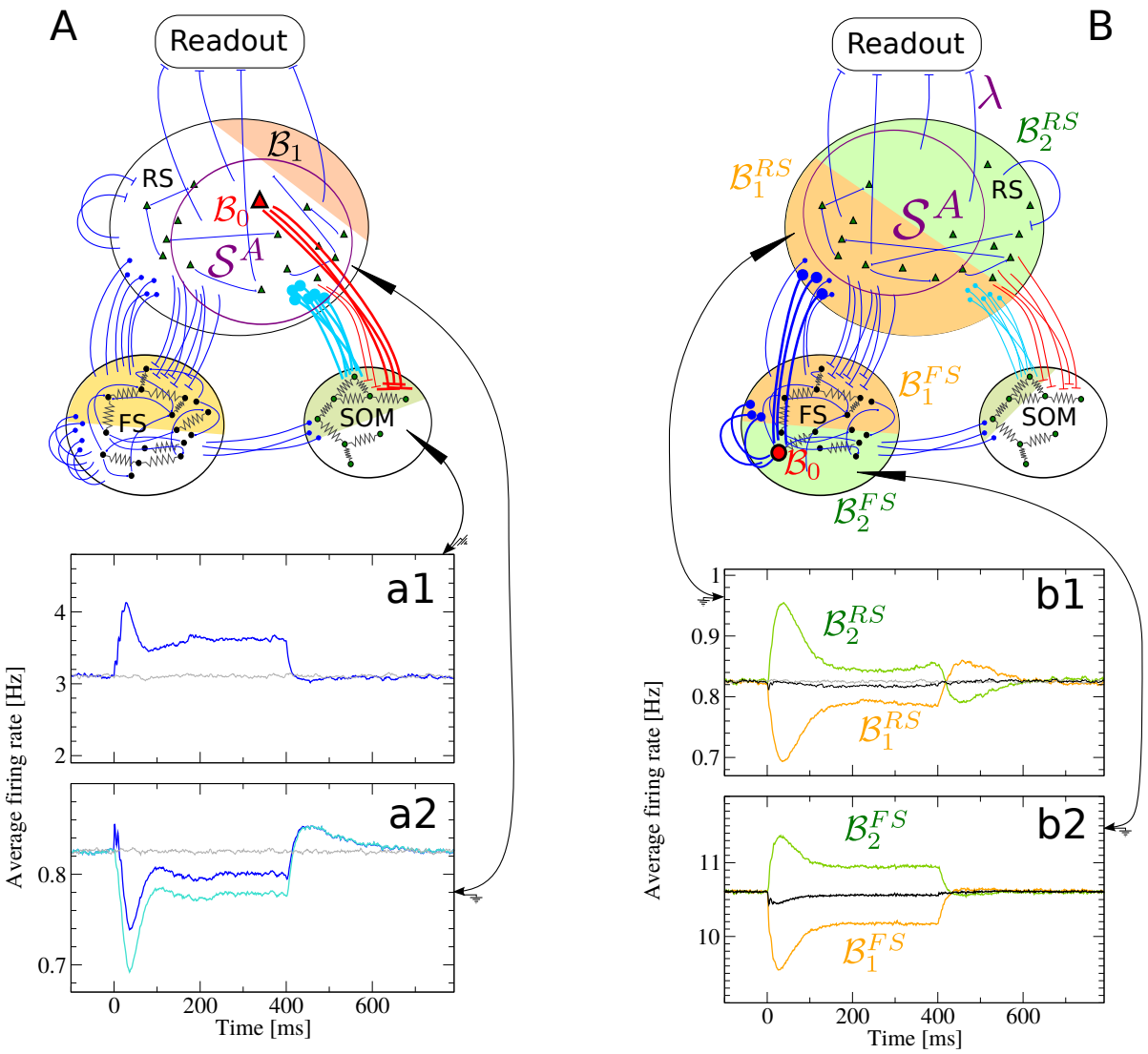

Figure 4.5. – **Trial-averaged firing-rate response to the stimulation of a RS and of a FS cell.** **A**: when a RS cell is stimulated, synapses from the stimulated cell ($\mathcal{B}_0$) to the SOM-LTS population strongly facilitate and cause a large increase in the firing rate of the SOM population, which then relaxes back to a plateau because of the spike-frequency adaptation (**a1**). The inhibitory input from the SOM to the RS population produces a response in the RS cells which is almost a mirror image (**a2**, blue line). The initial positive peak in the RS response is due to the spikes fired by $\mathcal{B}_0$ itself, as it can be seen by excluding $\mathcal{B}_0$ (**a2**, light blue). **B**: when a FS cell is stimulated, the firing rate of RS cells receiving direct input from $\mathcal{B}_0$ ($\mathcal{B}_1^{RS}$) drops and then recovers to an intermediate value (**b1**, orange line) because of the short-term depression of FS-to-RS synapses. Other cells within the RS population ($\mathcal{B}_2^{RS}$) respond in an almost perfectly symmetrical way (**a2**, green line). This increase of the $\mathcal{B}_2^{RS}$ firing activity is due to the overall decrease in firing rate in the FS population, if $\mathcal{B}_0$ is excluded (**b2**, black line). This reduction results from averaging the firing rate of the subpopulation $\mathcal{B}_1^{FS}$, which is directly inhibited (**b2**, orange line), and the firing rate of $\mathcal{B}_2^{FS}$, which is disinhibited (**b2**, green line). All curves are based on averages over 10000 trials. Grey lines represent catch trials.

receive direct inhibition from $\mathcal{B}_0$ (indicated by $\mathcal{B}_2^{FS}$) increase their firing rate (fig. 4.5b2, green line). However, the overall effect on the FS population is inhibitory, as shown by the black line in fig. 4.5b2, which is the firing rate of the entire FS population *excluding* $\mathcal{B}_0$. This is the set of cells that provide inhibitory input to $\mathcal{B}_2^{RS}$, i.e. RS neurons that do not receive direct input from $\mathcal{B}_0$, which therefore *increase* their firing rate. As it can be seen in fig. 4.5b1 (green line), the average firing rate of $\mathcal{B}_2^{RS}$ is almost a mirror image around the spontaneous level of the firing rate of $\mathcal{B}_1$. Consequently, when the average over the entire RS population is considered (the average size of $\mathcal{B}_1$ and of $\mathcal{B}_2$ is the same), the two opposite responses cancel almost completely (fig. 4.5b2, black line). Therefore, the stimulation of a FS cell is very hard to detect if the readout is not biased, as in the previous chapters. If λ indicates the fraction of cells chosen from the $\mathcal{B}_1^{RS}$ population, the natural value of λ, i.e. that corresponding to no readout bias, is $\lambda = \lambda_0 = 0.5$. In the following, two possible bias values will be considered: $\lambda = \lambda_+ = 0.05$ and $\lambda = \lambda_- = 0.9$, which correspond to a quite strong bias against and towards $\mathcal{B}_1^{RS}$, respectively.

4.1.4. Readout

In this section, the three possible readout mechanisms are described in detail. The first readout is the integrator readout, similar to the readout scheme A of the previous chapter.

Integrator readout

A random selection of $\hat{C} = 1000$ RS neurons constitutes the readout set $\mathcal{S}^A$. As opposed to the previous chapters, the stimulated cell ($\mathcal{B}_0$) is not treated differently. Therefore, when $\mathcal{B}_0$ is a RS cell, it can be randomly selected as part of $\mathcal{S}^A$ (because the readout cells are chosen from RS, an inhibitory FS $\mathcal{B}_0$ cannot be part of the readout set). The previous section showed how the response of the RS population is quite different depending on whether $\mathcal{B}_0$ is a RS or a FS neuron. More precisely, the response to the stimulation of a RS neuron is mediated by the SOM population and affects the RS population quite homogeneously. Hence, it does not depend strongly on the particular choice of neurons. On the contrary, when $\mathcal{B}_0$ is a FS cell, RS neurons respond differently, depending on whether they are chosen from $\mathcal{B}_1$ or from $\mathcal{B}_2$. If neurons are chosen at random to form the readout set, neurons from $\mathcal{B}_1$ and $\mathcal{B}_2$ are equally probable, and, as made clear in the discussion of the previous section (see also the black line in fig. 4.5b2), the average response is very faint. Therefore, when the stimulated cell is a FS neuron, a readout bias will be used, as in the previous chapters, whereas no readout bias is necessary in the case that $\mathcal{B}_0$ is a RS neuron. Because the focus of the present chapter is the dependence of the detectability on diverse stimulus properties, the readout bias will be left constant at either $\lambda = \lambda_+ = 0.05$ or $\lambda = \lambda_- = 0.9$.

The spike trains emitted by all neurons within the readout population $\mathcal{S}^A$ are then filtered, as in the previous chapters. However, instead of employing a filter of fixed shape, here a dynamical

equation is used to obtain the readout activity $v_{\rm ir}(t)$:

$$\tau_{m,e}\dot{v}_{\rm ir}(t) = -v_{\rm ir} + R_{m,read}\left[\sum_{i\in\mathcal{S}^A} J_{read,i}(t)x_i(t)\right], \tag{4.21}$$

where x_i is the spike train of the ith neuron within the readout set $\mathcal{S}^A$, the integration time constant is taken equal to the membrane time constant of RS neurons, and $R_{m,read} = \tau_{m,e}/C_m$. The dynamic weights $J_{read,i}(t)$ obey the same equation as all excitatory weights within the BCN:

$$J_{read,i}(t) = J_{ee}^{FF} R_i(t^-), \tag{4.22}$$

where the time-evolution of the depression variable $R_i(t)$ is governed by eq. (4.12) with parameters for strong depression ($\tau_{D,s} = 150\,\text{ms}$ and $U_{se,s} = 0.2$). Note that no fire-and-reset rule is applied to eq. (4.21), which is then equivalent to a linear filtering of the input spike trains with an exponential kernel with history-dependent amplitude.

To compute false positive and correct detection rates, a single boundary was used. Analogously to chapter 3, a detection event is registered if $v_{\rm ir}$ crosses at least once the boundary $\theta_\pm$ within a detection window, which is here $(0, T_w)$. As mentioned above, catch trials were simulated, in which no stimulus was present. These trials were used to determine the *false positive rate*. If a single lower detection boundary θ_- is used, the false positive rate is

$$\mathcal{FP}_{\rm ir}(\theta_-) = \left\langle \max_{t\in(0,T_w)} \Big\{ H(\theta_- v_{ir}(t)) \Big\} \,\middle|\, \text{no stimulation} \right\rangle. \tag{4.23}$$

The *hit* or *correct detection* rate is computed exactly in the same way, but in the presence of a stimulus

$$\mathcal{CD}_{\rm ir}(\theta_-) = \left\langle \max_{t\in(0,T_w)} \Big\{ H(\theta_- v_{ir}(t)) \Big\} \,\middle|\, \text{stimulation} \right\rangle. \tag{4.24}$$

The *effect size* as a function of the threshold is defined as in the previous chapters

$$\mathcal{Y}_{\rm ir}(\theta_-) = \mathcal{CD}_{\rm ir}(\theta_-) - \mathcal{FP}_{\rm ir}(\theta_-). \tag{4.25}$$

When an upper detection boundary is used, the false positive rate is

$$\mathcal{FP}_{\rm ir}(\theta_+) = \left\langle \max_{t\in(0,T_w)} \Big\{ H(v_{ir}(t) - \theta_+) \Big\} \,\middle|\, \text{no stimulation} \right\rangle \tag{4.26}$$

Correct detection rate and effect size are computed analogously.

Differentiator readout

The differentiator readout (DR) reads in the input from the network in the same was as the IR. In fact, it considers the difference between $v_{\rm ir}$ evaluated at two time points separated by a lag ΔT and then it convolves it with a smoothing filter $\mathcal{F}_{\tau_f}(t)$ to reduce the noise

$$v_{\rm dr}(t) = \Big(v_{\rm ir}(t) - v_{\rm ir}(t-\Delta T)\Big) \star \mathcal{F}_{\tau_f}(t). \tag{4.27}$$

The filter has an exponential shape, which can be viewed as a leaky integration:

$$\mathcal{F}_{\tau_f}(t) = \frac{e^{-t/\tau_f}}{\tau_f}. \tag{4.28}$$

Trajectories computed from eq. (4.27) are used in combination with an upper detection threshold θ^+ to obtain the false positive and hit rates, similarly as before:

$$\mathcal{FP}_{\rm dr}(\theta_+) = \left\langle \max_{t\in(0,T_w)} \Big\{ H(v_{dr}(t)-\theta_+) \Big\} \,\middle|\, \text{no stimulation} \right\rangle \tag{4.29}$$

$$\mathcal{CD}_{\rm dr}(\theta_+) = \left\langle \max_{t\in(0,T_w)} \Big\{ H(v_{dr}(t)-\theta_+) \Big\} \,\middle|\, \text{stimulation} \right\rangle \tag{4.30}$$

$$\mathcal{Y}_{\rm dr}(\theta_+) = \mathcal{CD}_{\rm dr}(\theta_+) - \mathcal{FP}_{\rm dr}(\theta_+). \tag{4.31}$$

Differentiator network readout

The operation performed by the DR, i.e. the subtraction of a delayed copy of the readout activity, can be approximately implemented by the readout architecture considered in section 3.3, provided that some details are modified accordingly. The differentiator readout network (DNR) consists, as in chapter 3, of two populations: one readout population of $N_B = 10\,000$ RS neurons ($\mathcal{S}^B$) and one population of $N_{\mathcal{I}} = 2000$ FS inhibitory neurons ($\mathcal{I}$). As in the previous chapter, both populations receive the same number of excitatory feed-forward connections from the RS population of the BCN. In this chapter, the number of feed-forward connections per neuron is $\hat{C} = 1000$. Neurons in the readout population $\mathcal{S}^B$ evolve according to the same dynamical equation as RS neurons of the BCN:

$$\tau_{m,k}\dot{v}_k = R_{m,k}\left[I_{{\rm ext},k}(t) + I_{{\rm rec},k}(t) + I_{\rm ff,k}(t) - a_k(t)\right]. \tag{4.32}$$

Neurons in the FS readout population, $\mathcal{I}$, obey

$$\tau_{m,k}\dot{v}_k = R_{m,k}[I_{{\rm ext},k}(t) + I_{{\rm rec},k}(t) + I_{\rm ff,k}(t)], \tag{4.33}$$

where the usual fire-and-reset rule applies to both last equations and $a_k(t)$ evolves according to eq. (4.4). All parameters relative to eqs. (4.32) and (4.33) are drawn from the same distributions used for RS and FS neurons within the BCN, respectively. The form of the external input term $I_{\text{ext},k}(t)$ is the same as in the BCN:

$$R_{m,k}I_{\text{ext},k}(t) = R_{m,k}I_0 + \tau_{m,k}\left[\sum_{j=1}^{C^R_{\text{ext,th}}}\sum_{l} J_{k,j,l}\delta(t-t_{k,j,l}) + \sum_{p=1}^{C^R_{\text{ext,bc,k}}}\sum_{q} J_{k,p,q}\delta(t-t_{k,p,q})\right]. \quad (4.34)$$

The firing rate of each input Poissonian spike train is the same as for the BCN, that is, $r_{\text{ext,th}} = 10\,\text{Hz}$ for the second term in eq. (4.34), and $r_{\text{bc,th}} = 2\,\text{Hz}$ for the third term. However, the number of "thalamic" inputs is smaller $C^R_{\text{ext,th}} = 250$, which is compensated by the constant term is $R_{m,k}I_0 = 15\,\text{mV}$. The number of "cortical" Poissonian inputs is $C^R_{\text{ext,bc,e}} = 2000$ for neurons within $\mathcal{S}^B$ and $C^R_{\text{ext,bc,i}} = 1000$ for neurons within $\mathcal{I}$.

The feed-forward input from the BCN to each neuron of the DNR has the same form for all neurons:

$$R_{m,k}I_{\text{ff},k}(t) = \tau_{m,k}\sum_{j\in\mathcal{Q}_e^\lambda(k)} J_{kj}(t-D_{kj})x_j(t-D_{kj}), \quad (4.35)$$

where $\mathcal{Q}_e^\lambda(k)$ is a set of $\hat{C}$ randomly selected neurons within the RS population of the BCN. As for the DR, when the stimulated cell is a FS neuron, neurons forming $\mathcal{Q}_e^\lambda(k)$ are chosen with probability $\lambda = \lambda_\pm$ from $\mathcal{B}_1$, whereas no bias is applied in the case that $\mathcal{B}_0$ is a RS neuron. Each weight $J_{kj}(t)$ obeys the usual equation for strong STD (with the same parameters as in the other readout variants)

$$J_{kj}(t) = J_{kj}R_j(t^-), \quad (4.36)$$

where the constant factor J_{kj} is drawn from an exponential distribution with mean $J^{FF}_{ee} = J_{ee} = 0.1\,\text{mV}$ when k is part of the $\mathcal{S}^B$ population, and from an exponential distribution with mean $J^{FF}_{ie} = J_{ie} = 0.2\,\text{mV}$ when k belongs to $\mathcal{I}$. Transmission delays are drawn from a uniform distribution in the interval 0.5 ms to 1 ms when k belongs to the RS readout population $\mathcal{S}^B$. When k belongs to $\mathcal{I}$, the delay is drawn from a uniform distribution of equal width (0.5 ms), but shifted by $\Delta T = 10\,\text{ms}$ (i.e. delays are uniformly distribution in the range 10.5 ms to 11 ms). Helmstaedter et al. (2008) measured latencies between action potentials and EPSP onsets in the rat barrel cortex and found an approximately linear relationship between inter-somatic distance and latency with an intercept of $\approx 0.5\,\text{ms}$ and a slope of $\approx 5\,\text{ms/mm}$. Therefore, a latency of $\approx 10\,\text{ms}$ would correspond to an inter-somatic distance of $\approx 2\,\text{mm}$ (the approximate diameter of a typical barrel is $\approx 0.3\,\text{mm}$).

Finally, the recurrent input term is

$$R_{m,k} I_{\mathrm{rec},k}(t) = -\tau_{m,k} \sum_{\ell \in \mathcal{L}_i(k)} J_{k\ell}(t - D_{k\ell}) x_\ell(t - D_{k\ell}), \tag{4.37}$$

where $\mathcal{L}_i(k)$ are sets of $C^R_{ii} = 200$ neurons selected at random from $\mathcal{I}$. Transmission delays in eq. (4.37) are drawn independently from a uniform distribution in the interval 0.5 ms to 1.0 ms, while the coupling weights $J_{k\ell}(t)$ follow the same STD dynamics as feed-forward connections and as most weights of connections within the BCN. The prefactor of each synapse is drawn independently from an exponential distribution. The mean of the weight distribution is $J^R_{ii} = J_{ii} = 1\,\mathrm{mV}$ for recurrent connections from $\mathcal{I}$ to $\mathcal{I}$, and J^R_{ei} for connections from $\mathcal{I}$ to $\mathcal{S}^B$. If the DNR is to implement the operation performed by the DR, the feed-forward inhibition from $\mathcal{I}$ to $\mathcal{S}^B$ must cancel the feed-forward excitatory input that the readout population $\mathcal{S}^B$ receives from the BCN.

To find an estimate of the value of J^R_{ei} that realizes this condition, consider the construction depicted in fig. 4.6. Suppose that the firing rate of the RS in the BCN changes by Δr_e, and that all time dependencies can be neglected. As a consequence of the change in the input firing rate, the mean input from the BCN to $\mathcal{S}^B$ changes by $\Delta\mu_e$, while the input from $\mathcal{I}$ to $\mathcal{S}^B$ changes by $\Delta\mu_I$. The change in the mean input from the BCN to $\mathcal{S}^B$ is

$$\Delta\mu_e = \tau_{m,e} J^{\mathrm{FF}}_{ee} \bar{R}(r_e) \hat{C} \Delta r_e, \tag{4.38}$$

where the term

$$\bar{R}(r) = \frac{1}{1 + \tau_{D,s} U_{se,s} r} \tag{4.39}$$

represents the average effect of the STD, given a presynaptic firing rate r. The mean input from $\mathcal{I}$ to $\mathcal{S}^B$ changes by an amount

$$\Delta\mu_{\mathcal{I}} = -\tau_{m,e} J^{\mathrm{R}}_{ei} \bar{R}(r_{\mathcal{I}}) C^R_{ei} \Delta r_{\mathcal{I}}, \tag{4.40}$$

which depends on $\Delta r_{\mathcal{I}}$, the change in the firing rate of $\mathcal{I}$ from the spontaneous value $r_{\mathcal{I}}$. The same arguments used in section 3.3 lead to the linear-order approximation to $\Delta r_{\mathcal{I}}$:

$$\Delta r_{\mathcal{I}} = \chi(0) \left(\tau_{m,i} J^{\mathrm{FF}}_{ie} \bar{R}(r_e) \hat{C} \Delta r_e - \tau_{m,i} J^{\mathrm{R}}_{ii} \bar{R}(r_{\mathcal{I}}) C^R_{ii} \Delta r_{\mathcal{I}} \right), \tag{4.41}$$

where $\chi(0) = \mathrm{d}\phi_{sn}/\mathrm{d}\mu$ is, as in the previous chapters, the DC susceptibility of the firing rate (see section 1.4.2). Equation (4.41) can be solved for $\Delta r_{\mathcal{I}}$ and substituted into eq. (4.40). Imposing

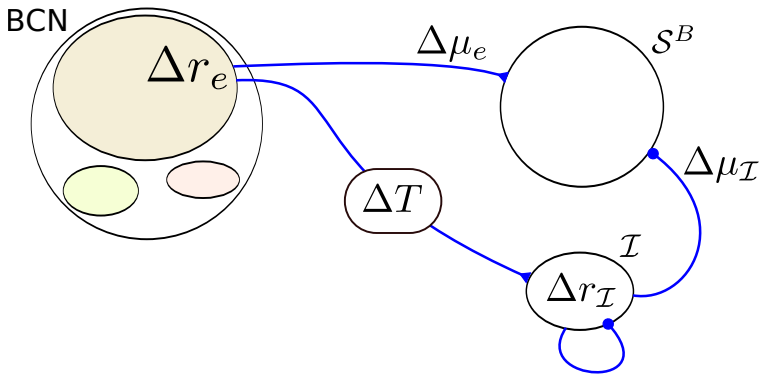

Figure 4.6. – **Tuning of the differentiator readout network to implement the operation of the differentiator readout scheme.** A perturbation in the firing rate of the RS neurons in the BCN (Δr_e) causes a perturbation in the mean input to the RS readout neurons, $\Delta\mu_e$, and a perturbation in the firing rate of the inhibitory readout population $\mathcal{I}$. This change in firing rate causes a shift in the input from $\mathcal{I}$ to $\mathcal{S}^B$ ($\Delta\mu_\mathcal{I}$). The strength of the connection from $\mathcal{I}$ to $\mathcal{S}^B$ is adjusted such that $\Delta\mu_e + \Delta\mu_\mathcal{I} = 0$. This cancellation reaches $\mathcal{S}^B$ with a time lag ΔT.

$\Delta\mu_e + \Delta\mu_\mathcal{I} = 0$ and solving for J^R_{ei} yields

$$J^R_{ei} = \frac{J^{FF}_{ee}\left(1 + \tau_{m,i}\chi(0)J^R_{ii}\bar{R}(r_\mathcal{I})C^R_{ii}\right)}{\tau_{m,i}\chi(0)J^{FF}_{ie}C^R_{ei}\bar{R}(r_\mathcal{I})}. \tag{4.42}$$

The spontaneous firing rate of $\mathcal{I}$ can be estimated from the numerical solution of the following self-consistency condition, analogous to eq. (2.10):

$$r_\mathcal{I} = \phi_{sn}(J^{FF}_{ee}, J^R_{ii}, r^{in}_{tot}, C^R_{ii}\cdot r_\mathcal{I}, I_{\text{ext}}), \tag{4.43}$$

where $r^{in}_{tot} = \hat{C}r_e + C^R_{\text{ext,bc,e}}r_{\text{ext,bc}} + C^R_{\text{ext,th,e}}r_{\text{ext,th}}$ is the total excitatory input rate to $\mathcal{I}$ and ϕ_{sn} is given by eq. (1.27). By substituting numerical values in eq. (4.42), one finds that $J^R_{ei} = 0.65\,\text{mV}$ approximately satisfies the imposed condition. With this choice of parameters, the spontaneous activity of the DNR is asynchronous irregular with firing rates $r_B \approx 1\,\text{Hz}$ $r_\mathcal{I} \approx 9\,\text{Hz}$.

The DNR activity is obtained by filtering the average firing rate of the readout neurons $\mathcal{S}^B$ with the same exponential filter used for the DR:

$$v_{\text{dnr}}(t) = \frac{1}{N_B}\sum_{x_k\in\mathcal{S}^B} x_k(t) \star \mathcal{F}_{\tau_f}(t), \tag{4.44}$$

where the filter $\mathcal{F}_{\tau_f}(t)$ is given by eq. (4.28). False positive and hit rates, and the effect size are

obtained in exactly the same way as done for the DNR:

$$\mathcal{FP}_{\mathrm{dnr}}(\theta_+) = \left\langle \max_{t\in(0,T_w)} \left\{ H(v_{dnr}(t) - \theta_+) \right\} \middle| \text{ no stimulation} \right\rangle, \tag{4.45}$$

$$\mathcal{CD}_{\mathrm{dnr}}(\theta_+) = \left\langle \max_{t\in(0,T_w)} \left\{ H(v_{dnr}(t) - \theta_+) \right\} \middle| \text{ stimulation} \right\rangle, \tag{4.46}$$

$$\mathcal{Y}_{\mathrm{dnr}}(\theta_+) = \mathcal{CD}_{\mathrm{dnr}}(\theta_+) - \mathcal{FP}_{\mathrm{dnr}}(\theta_+). \tag{4.47}$$

As in the previous chapters, the false positive rate of 0.25 (which corresponds approximately to the average false positive rate measured experimentally) was chosen to compare the simulation results to the experimental data. More precisely, the threshold $\bar{\theta}$ is chosen such that

$$\mathcal{FP}_X(\bar{\theta}) = 0.25, \tag{4.48}$$

which is then used to compute

$$\bar{\mathcal{Y}}_X = \mathcal{Y}_X(\bar{\theta}). \tag{4.49}$$

In the last two equations X indicates the detector type and θ can be either an upper or lower boundary.

4.2. Results

The purpose of this section is to use the network model and the three readout schemes introduced in the last section to investigate how the detectability of the single-cell stimulation depends on the properties of the injected current.

The stimuli used here replicate exactly those used experimentally and are divided in three sets, as shown in fig. 4.7. The three stimuli in fig. 4.7A have equal intensity but different length, and in the experiment they were used to study how the effect size depends on the number of evoked spike trains. The intensity of the three stimuli shown in fig. 4.7B is chosen such that their area is constant. In the experiment, they were employed to investigate how the effect size depends on the firing rate of the evoked spike train while keeping the number of spikes approximately constant. Finally, randomly shuffled sequences of six current steps different in height and duration were used to elicit irregular spike trains (fig. 4.7C). In this final experiment, the effect of the regularity of the evoked spike train (quantified by the CV of the elicited spike train) on the detectability was examined.

The central quantity characterizing the detectability of the single-cell stimulation will be the

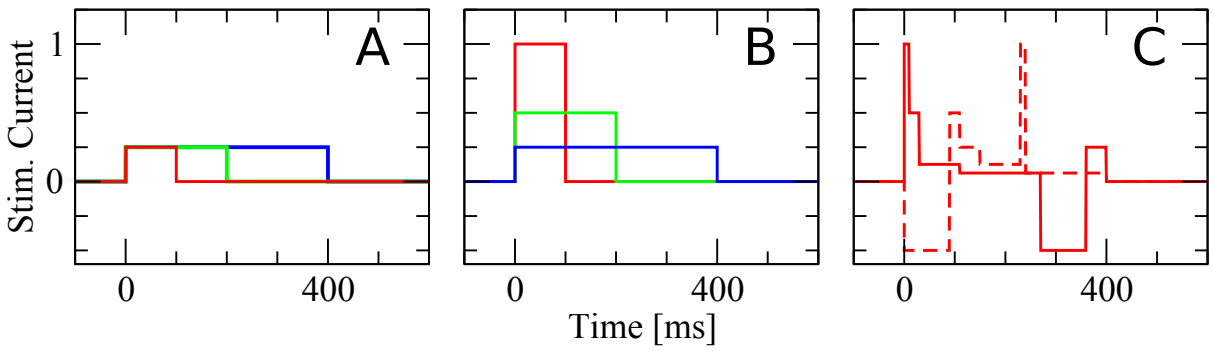

Figure 4.7. – **Types of stimuli used in this chapter**. **A**: stimuli have the same intensity and different duration. **B**: stimuli have intensity inversely proportional to the duration. **C**: stimuli are a random permutation of five positive and one negative current steps with different intensity and duration. In all cases, the current intensity is normalized to the maximum current (reported in table 4.5).

effect size, which will be considered as a function of the false positive rate:

$$\mathcal{Y}_X(\mathcal{FP}) = \mathcal{CD}_X(\mathcal{FP}) - \mathcal{FP}_X, \tag{4.50}$$

where X indicates the detector type (X = IR, DR, DNR). Equation (4.50) is equivalent to the receiver operating characteristic (ROC) curve of the detector (see fig. 2.6 in section 2.3.3) minus the diagonal. As in the previous chapters, a false positive rate of 25% (approximately the average false positive rate measured in the experiments) is chosen for the direct comparison with the experimental data:

$$\bar{\mathcal{Y}}_X = \mathcal{CD}_X\left(\frac{1}{4}\right) - \frac{1}{4}. \tag{4.51}$$

Besides the effect size, two further quantities will be measured and discussed, namely the mean and standard deviation of the readout activity. Statistics of higher order (skewness and kurtosis) do not display appreciable deviations from the spontaneous state and will be omitted for brevity. Mean and standard deviation of the readout activity will be plotted as standardized deviations from the spontaneous value. More precisely, considering first the time-dependent mean:

$$\hat{\mu}_X(t) = \frac{\left\langle v_X(t)\right\rangle - \mu_{X,\,\mathrm{catch}}}{\sigma_{X,\,\mathrm{catch}}}, \tag{4.52}$$

where $\mu_{X,\,\mathrm{catch}}$ and $\sigma_{X,\,\mathrm{catch}}$ are the average mean and standard deviation in the spontaneous state, respectively:

$$\mu_{X,\,\mathrm{catch}} = \left\langle v_X(t) \,\middle|\, \text{no stimulation}\right\rangle \tag{4.53}$$

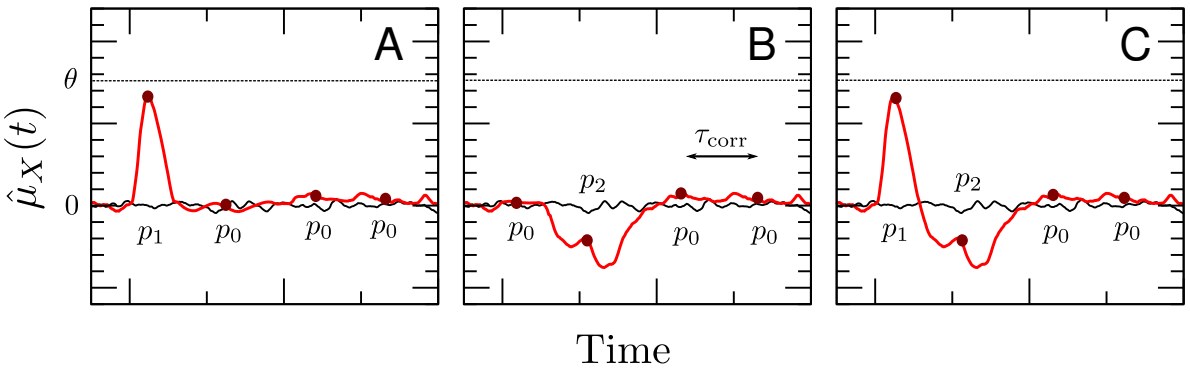

Figure 4.8. – **Illustration of simplified detection model**

$$\sigma_{X,\,\text{catch}} = \left\langle \sqrt{\Delta v_X^2(t)} \,\middle|\, \text{no stimulation} \right\rangle, \tag{4.54}$$

where $\Delta v_X^2(t) = \Big(v_X(t) - \big\langle v_X(t) \big\rangle\Big)^2$, and the time dependence in both last equations is self-averaging due to the stationary conditions. The time-dependent standard deviation of the readout activity is standardized in the same way:

$$\hat{\sigma}_X(t) = \frac{\left\langle \sqrt{\Delta v_X^2(t)} \right\rangle - \sigma_{X,\,\text{catch}}}{\sigma_{X,\,\text{catch}}}. \tag{4.55}$$

Any non-zero values of $\hat{\mu}_X(t)$ and of $\hat{\sigma}_X(t)$ influence the effect size in different ways. Suppose, for concreteness, that the detector uses an upper boundary. In this case, a positive deflection of $\hat{\mu}_X(t)$ increases locally the probability of reaching the decision threshold θ_+, while a negative deflection reduces it. If a lower detection boundary is used, the opposite holds. Non-zero values of $\hat{\sigma}_X(t)$ have the same effect independently of the kind of barrier used: a local increase in the standard deviation always enhances the probability of a threshold crossing, whereas a negative value of $\hat{\sigma}_X(t)$ always decreases the probability of reaching the barrier.

To understand how multiple deviations from the spontaneous state can combinedly affect the effect size, it is useful to consider a simplified detection model analogous to the detection theory developed in section 2.3.3. In this theory, detection rates are approximated as the result of $n = T_w/\tau_{\text{corr}}$ draws of a discrete (Gaussian) variable, where T_w is the detection time window and τ_{corr} is the autocorrelation time of the readout variance (in the simplified example of fig. 4.8 $n = 4$). In particular, the false positive rate is given by

$$\mathcal{FP}(\theta) = 1 - p_0^n(\theta), \tag{4.56}$$

where $p_0(\theta)$ is the probability of *not* crossing the barrier θ at any given time point. Let us set, for concreteness, the barrier at the value $\bar{\theta}$, which gives the standard false positive rate of $1/4$, so that the dependence on θ can be dropped (but the derivation below is valid for any value of θ). Suppose now that one ensemble of trajectories exhibits one feature increasing the probability of triggering the detector, such as a positive peak in $\hat{\mu}_X(t)$ when an upper barrier is used (fig. 4.8A). In the vicinity of the peak, the probability of *not* triggering the detector will be $p_1 = p_0 + \Delta p_1 < p_0$. The correct detection rate for this situation is

$$\bar{\mathcal{CD}}_1 = 1 - p_1 p_0^{n-1}, \tag{4.57}$$

so that the effect size is

$$\bar{\mathcal{Y}}_1 = 1 - p_1 p_0^{n-1} - (1 - p_0^n) = p_0^n \left(1 - \frac{p_1}{p_0}\right). \tag{4.58}$$

Suppose now that another feature decreases locally the probability of reaching the threshold, such as a negative deflection in $\hat{\mu}_X(t)$ (as in fig. 4.8B), or a decrease in the standard deviation. Locally, the probability of not triggering the detector will be $p_2 = p_0 + \Delta p_2 > p_0$ and the effect size for such a scenario is

$$\bar{\mathcal{Y}}_2 = 1 - p_2 p_0^{n-1} - (1 - p_0^n) = p_0^n \left(1 - \frac{p_2}{p_0}\right). \tag{4.59}$$

Suppose now that both features are present at sufficiently separated times within the same detection time window, as in fig. 4.8C. In this case, the effect size is

$$\bar{\mathcal{Y}}_{12} = p_0^n - p_1 p_2 p_0^{n-2} = p_0^n \left(1 - \frac{p_1 p_2}{p_0^2}\right). \tag{4.60}$$

Substituting $p_1 = p_0 + \Delta p_1$ and $p_2 = p_0 + \Delta p_2$ into eq. (4.60) and supposing $\Delta p_1, \Delta p_2 \ll 1$ yields

$$\begin{aligned} \bar{\mathcal{Y}}_{12} &= p_0^n \left(1 - \frac{p_0^2 + p_0\Delta p_1 + p_0\Delta p_2 + \Delta p_1 \Delta p_2 + p_0^2 - p_0^2}{p_0^2}\right) \\ &\approx p_0^n \left(1 - \frac{p_0(p_0 + \Delta p_1)}{p_0^2} + 1 - \frac{p_0(p_0 + \Delta p_2)}{p_0^2}\right) \\ &= \bar{\mathcal{Y}}_1 + \bar{\mathcal{Y}}_2. \end{aligned} \tag{4.61}$$

This approach can be generalized to the case of more deviations from the spontaneous state. For instance, when three features are present

$$\bar{\mathcal{Y}}_{123} = p_0^n \left(1 - \frac{p_1 p_2 p_3}{p_0^3}\right) \tag{4.62}$$

can also be similarly expanded by inserting $p_i = p_0 + \Delta p_i$ (where $i = 1, 2, 3$) into eq. (4.62) and neglecting all terms that contain products of at least two of the three Δp_i:

$$\begin{aligned}\bar{\mathcal{Y}}_{123} &\approx p_0^n \left(1 - \frac{p_0^3 + p_0^2\Delta p_1 + p_0^2\Delta p_2 + p_0^2\Delta p_3 + 2p_0^3 - 2p_0^3}{p_0^3}\right) \\ &= p_0^n \left(1 - \frac{p_0^2(p_0 + \Delta p_1)}{p_0^3} + 1 - \frac{p_0^2(p_0 + \Delta p_2)}{p_0^3} + 1 - \frac{p_0^2(p_0 + \Delta p_3)}{p_0^3}\right) \\ &= \bar{\mathcal{Y}}_1 + \bar{\mathcal{Y}}_2 + \bar{\mathcal{Y}}_3.\end{aligned} \tag{4.63}$$

This toy model shows that favorable and unfavorable deviations from the spontaneous state appearing in the same detection window can (approximately) add or cancel each other, which will be useful to interpret the influence of $\hat{\mu}_X(t)$ and $\hat{\sigma}_X(t)$ on the effect size.

4.2.1. Effect of stimulus duration

The effect of changing the stimulus duration will be considered first. To this end, stimuli of length 100, 200, and 400 ms are used. The stimulus intensity is kept constant at 25% of the maximum current. In the experiment, when the stimulated cell was a RS neuron, the three stimuli evoked 6 ± 3, 11 ± 5, and 23 ± 10 spikes, respectively. In the model, the number of evoked spikes was similar, being 7 ± 1, 12 ± 2, and 20 ± 5 spikes. In the case that a FS cell was stimulated, the experimentally observed number of evoked spikes during current injection was 13 ± 7, 30 ± 14, and 61 ± 38 spikes, for the 100, 200, and 400 ms stimulus, respectively. In the model, the same stimuli elicited 12 ± 2, 24 ± 3, and 47 ± 7 spikes, respectively. For both cell types, the average number of evoked spikes generated by the model is well within one standard deviation of the experimental data. However, the spread of the spike count distribution is smaller in the model, which is not surprising, considering the multiple possible noise sources that are not modeled, and that only some of the cellular parameters are randomly distributed in the model.

Figure 4.9 gives an overview of $\hat{\mu}_X(t)$, $\hat{\sigma}_X(t)$, and the effect size measured by all three detectors in the case of the stimulation of a RS neuron. The plots are organized as follows. The first column refers to the integrator readout (IR), the second to the differentiator readout (DR), the third to the differentiator network readout (DNR). The first row shows the standardized difference from the spontaneous value of the time-dependent mean, $\hat{\mu}_X(t)$, the second row displays the standardized difference from the spontaneous value of the standard deviation of the readout activity, $\hat{\sigma}_X(t)$, and the third reports the effect size as a function of the false positive rate, eq. (4.50). In fig. 4.9, the color of each curve corresponds to one stimulus, as shown in fig. 4.7A: red refers to the 100 ms stimulus, green to the to the 200 ms stimulus, and blue to the 400 ms stimulus. The black thin line show results for catch trials.

The IR activity, defined in eq. (4.21), is a low-pass filtered sum of the spike trains of a

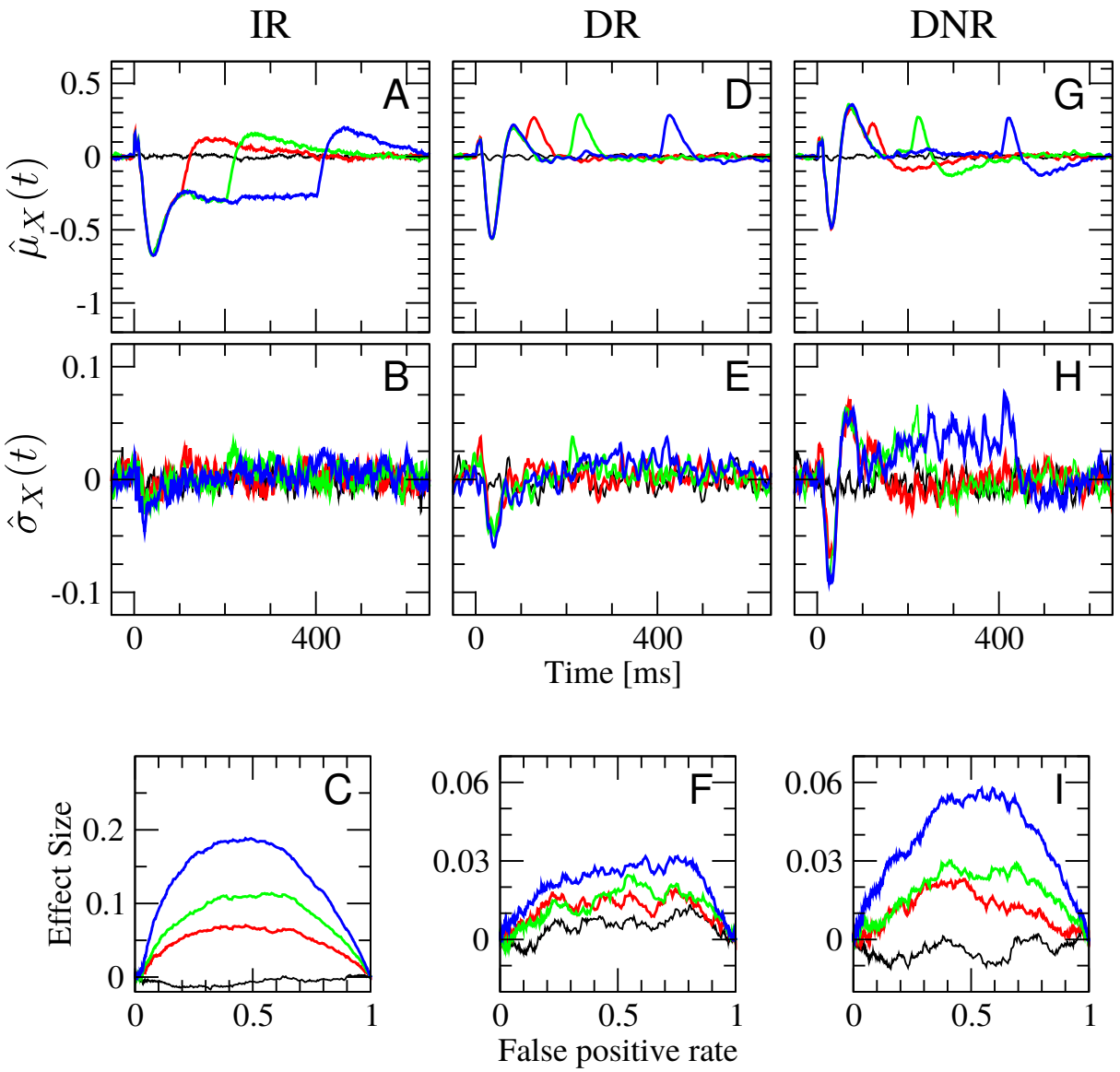

Figure 4.9. – **Summary of detection statistics upon stimulation of a RS cell with the three stimuli in fig. 4.7A (equal intensity, different duration).** First row: standardized deviation from the spontaneous value of the time-dependent mean readout activity eq. (4.52). Second row: standardized deviation from the spontaneous value of the time-dependent standard deviation of the readout activity eq. (4.55). Third row: effect size as a function of the false positive rate. First column: integrator readout (IR). Second column: differentiator readout (DR). Third column: differentiator network readout (DNR). The color of each lines corresponds to a stimulus as in fig. 4.7A. Black line is catch trial condition (no stimulus).

random subset of the RS population, where the amplitude of the filter keeps memory of the past activity. Therefore, it is not surprising that the time course of $\hat{\mu}_{\mathrm{ir}}(t)$ looks similar to the firing-rate response discussed in the previous section. As seen in fig. 4.9A, the initial response to the stimulus onset is a small positive peak, due to the spikes fired by $\mathcal{B}_0$ itself, followed by inhibition, caused by the activation of the SOM-LTS cells, (see also fig. 4.9). This inhibition peaks about 50 ms after the stimulus onset and then relaxes back to a plateau, because of the spike-frequency adaptation of SOM-LTS cells. The first 100 ms are identical for all three stimuli. When the stimulus is turned off, the mean IR activity quickly relaxes back to the spontaneous value and then overshoots. The overshoot is mostly due to spike-frequency adaptation in the RS cells: during the stimulation, RS cells have been firing below their spontaneous value, thus causing a decrease in the mean adaptation current, which, in turn, has a disinhibiting effect when the stimulus is turned off. Note that the overshoot is slightly larger for longer stimuli, because the adaptation variable has been "charging" for a longer time.

The deviations of $\hat{\sigma}_{\mathrm{ir}}(t)$ from the baseline level are rather modest in amplitude (fig. 4.9B) and their sign follows that of the mean: a dip can be seen right after the stimulus onset, while a small increase occurs coincidentally with the overshoot. The detectability of the three stimuli is quantified by the effect size as a function of the false positive rate, shown in fig. 4.9C and obtained by plotting eq. (4.25) as a function of eq. (4.23) upon variation of the lower detection boundary θ_-, as the main deviation from the spontaneous state is here in the negative direction. There is a clear difference in the detectability of the three stimuli: the longer the stimulus, the larger the effect size. If a FP level in the intermediate range is taken as reference, it can be seen that the effect size grows almost proportionally to the length of the stimulus. Although the distance of the plateau from the zero level is only ≈ 0.3 of the standard deviation, increasing its length provides an advantage to longer signals, which is in contrast to the experimental results.

The results obtained from the DR are quite different. The time-dependent deviation of the mean $\hat{\mu}_{\mathrm{dr}}(t)$ (fig. 4.9D) displays three positive peaks and one negative peak for all three signals, although the second and third peak partially overlap for the 100 ms signal. Each positive peak corresponds to an upswing of $\hat{\mu}_{\mathrm{ir}}(t)$, and the negative peak corresponds to the first downswing of the IR activity. The strongest deviation from the spontaneous state of the time-dependent standard deviation $\hat{\sigma}_{\mathrm{dr}}(t)$ (fig. 4.9E) is the dip in coincidence with the negative peak in the mean. The effect size as a function of the FP rate resulting from the DR activity is computed by combining eq. (4.29) and eq. (4.31) and shown in fig. 4.9F. It is much smaller than in fig. 4.9C and in the range of the experimentally measured average effect size (1 % to 2 %). Importantly, the differences between the three signals are minimal: the 100 ms and 200 ms signals are equally effective, and the 400 ms signal is only marginally better. Each signal causes an equal number of positive and negative deflections in $\hat{\mu}_{\mathrm{dr}}(t)$ and $\hat{\sigma}_{\mathrm{dr}}(t)$, which are also similar in amplitude. The 400 ms signal, however, induces a slight increase in the variance towards the end of the

stimulation. Taken together, these observation give a plausible explanation of the very similar effect size of the three signals, and of why the longest signal is still slightly easier to detect.

The time-dependent deviation of the mean DNR activity from the spontaneous value $\hat{\mu}_{\mathrm{dnr}}(t)$ (fig. 4.9G) has a qualitative shape that resembles that of the DR activity, thus supporting the conclusion that the DNR roughly operates as a differentiator circuit. However, the positive peaks have a slightly different shape and amplitude in comparison to those of fig. 4.9D. Another difference is the more pronounced undershoot after the last peak of each signal, which corresponds to the slow decay of the IR activity after the stimulus is switched off. The time-course of $\hat{\sigma}_{\mathrm{dnr}}(t)$ (fig. 4.9H) follows that of the mean, in that it shows positive and negative peaks roughly in the same position. However, these deviations from the spontaneous state are smaller in amplitude, and another difference can be seen, most evidently in the 400 ms signal: a weak but persistent increase between the initial and final peak. This increase in variance during the stimulation helps the trajectories to reach the barrier and leads to a larger effect size (fig. 4.9I) in comparison to the DR. Again, the longest signal is better than the other two because of the persistent increase in the readout standard deviation, which favors the long signal.

A direct comparison of the model results with the data is shown in fig. 4.10, which reports the average effect size measured from the data together with $\bar{\mathcal{Y}}_X$, the effect size corresponding to a false positive rate of 0.25, according to eq. (4.49). The three panels show results for the three detectors superimposed to the experimental results, which are the same for each panel. Figure 4.10A shows that the effect size computed from the IR, $\bar{\mathcal{Y}}_{\mathrm{ir}}$ (filled circles) overestimates the experimental data (open circles with error bars). Furthermore, it strongly depends on the stimulus type. However, $\bar{\mathcal{Y}}_{\mathrm{dr}}$, the effect size measured from the DR lies within the error bars of the experimental results and shows barely any dependence on the stimulus length (fig. 4.10B). The effect size measured from the DNR is similar to $\bar{\mathcal{Y}}_{\mathrm{dr}}$, although the effect size is slightly larger for the 400 ms stimulation (fig. 4.10C).

Altogether, the results presented above make clear that the DR provides the best agreement with the experimental findings, and that the DNR can approximately reproduce the operating principle of a differentiator, as far as the mean is concerned. The time-dependent standard deviation of the DNR activity shows some differences from that of the DR, which ultimately lead to a slight increase of the effect size if the stimulus length is increased. However, this effect is modest. The IR yields results in marked contrast with the data, because the effect size strongly depends on the duration of the injected current.

The rest of this section deals with the case that the stimulated cell is a FS neuron. Simulation results are shown in fig. 4.11 and are organized in the same way as in the case of the RS cell. The color coding of all lines is also the same, with one addition. Because here two values of the readout bias λ are considered, the entire plot should be repeated twice. Instead, the value $\lambda = \lambda_+ = 0.05$ will be considered as standard case, and the case of $\lambda = \lambda_- = 0.9$ will be shown

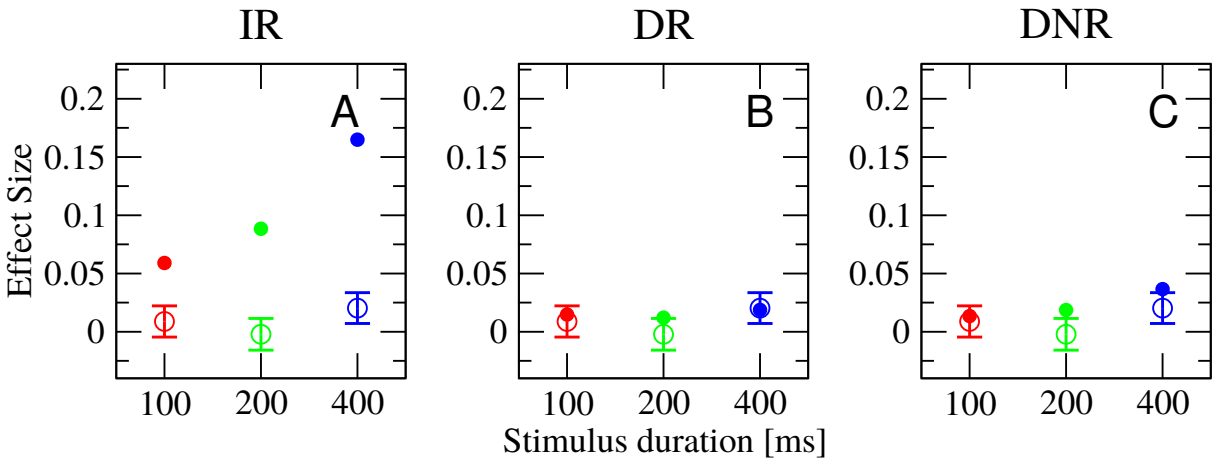

Figure 4.10. – **Results from differentiator readout (DR) and differentiator network readout (DNR) are compatible with experimental data, whereas integrator readout (IR) gives qualitatively different results (stimulation of RS cell, signals as in fig. 4.7A).** Open circles with error bars are experimental results, which are the same in each panel, and represent the average effect size computed from 119 RS cells. Experimental data are from the Brecht lab. The number of trials per cell is rather heterogeneous (total number of trials is 2407). **A**, **B**, and **C**: filled dots are the effect size (for a false positive rate of 0.25) resulting from the IR, DR, and DNR, respectively.

only for the longest stimulus as a blue dashed line.

The time course of $\hat{\mu}_{\mathrm{ir}}(t)$ for $\lambda = \lambda_+$ (fig. 4.11A, solid lines) shows an initial peak around 50 ms and a relaxation to an intermediate value owing to the short-term depression of the input synapses. When the stimulus is switched off, the readout activity undershoots and then finally relaxes back to the spontaneous value. As explained in section 4.1.3, when $\lambda = \lambda_+$ the readout population consists prevalently of neurons that do not receive direct input from $\mathcal{B}_0$, and that are disinhibited by FS neurons directly targeted by $\mathcal{B}_0$. However, when the bias is $\lambda = \lambda_-$, readout neurons are chosen prevalently from $\mathcal{B}_1$, i.e. neurons that are directly inhibited by $\mathcal{B}_0$ (fig. 4.11A, blue dashed line), and the response is essentially a mirror image. In both cases, the time course of $\hat{\sigma}_{\mathrm{ir}}(t)$ (fig. 4.11B) has the same shape as that of $\hat{\mu}_{\mathrm{ir}}(t)$. The effect size as a function of the FP rate (fig. 4.11C) shows that the largest contribution to the detectability is due to the initial peak, as the difference between the three curves is less pronounced than in the previous case. However, the later part of the response also gives a contribution, as the detectability clearly increases for longer stimuli. The effect size is larger in the case of $\lambda = \lambda_-$ (blue dashed line), probably because the distance of the plateau from zero is almost twice as large than in the case of $\lambda = \lambda_+$ (≈ -0.2 vs. ≈ 0.1).

The DR average readout activity (fig. 4.11D, continuous lines) displays a first pronounced positive peak corresponding to the upswing of the IR activity, followed by a milder but broader trough due to the relaxation to the plateau of the IR. When the stimulus is switched off, a

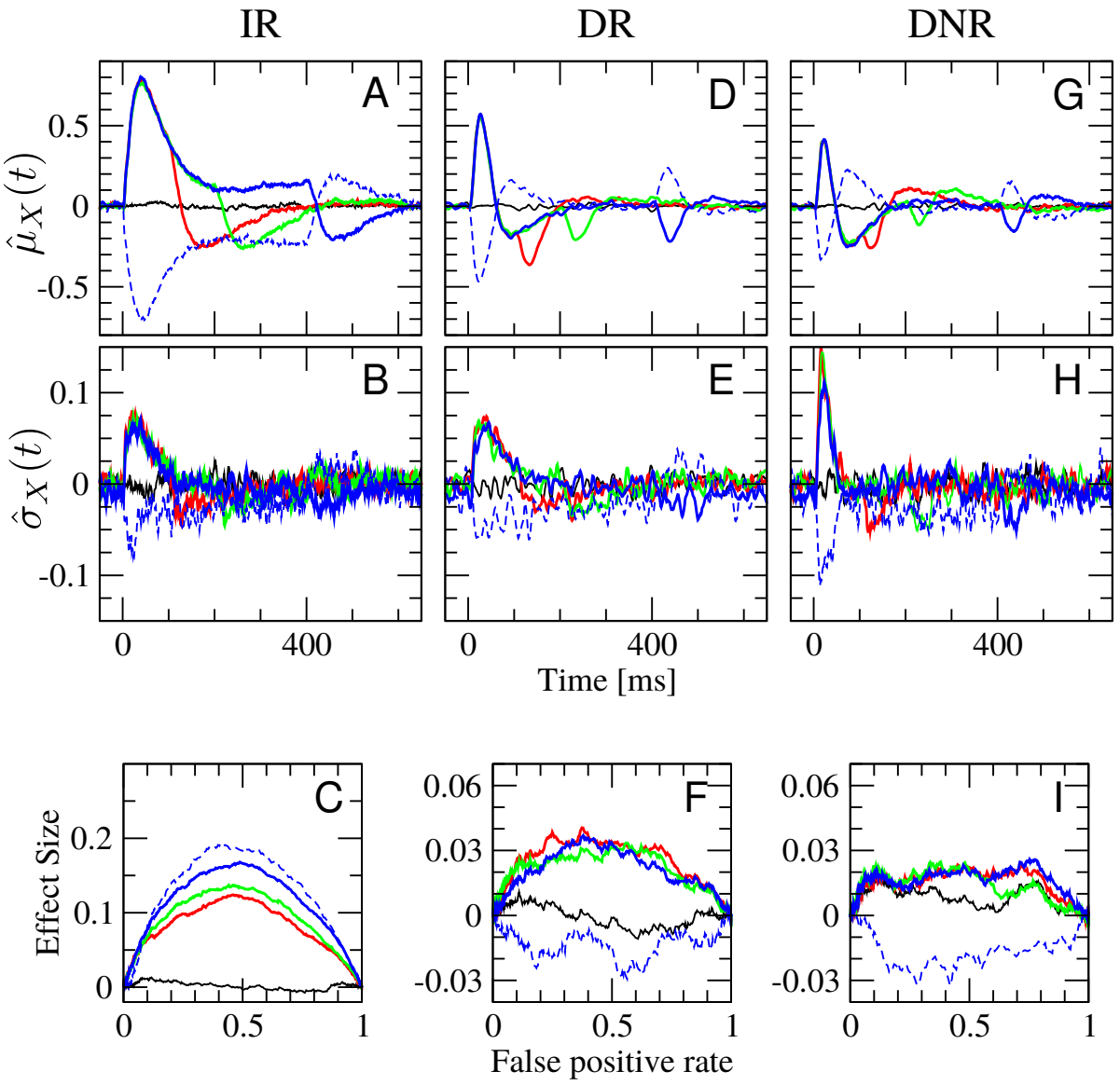

Figure 4.11. – **Summary of detection statistics upon stimulation of a FS cell with the three stimuli in fig. 4.7A (equal intensity different duration).** First row: standardized deviation from the spontaneous value of the time-dependent mean readout activity eq. (4.52). Second row: standardized deviation from the spontaneous value of the time-dependent standard deviation of the readout activity eq. (4.55). Third row: effect size as a function of the false positive rate. First column: integrator readout (IR). Second column: differentiator readout (DR). Third column: differentiator network readout (DNR). The color of each lines corresponds to a stimulus as in fig. 4.7A. Black line is catch trial condition (no stimulus). Blue dashed line refers to 400 ms stimulus with readout bias $\lambda_- = 0.9$. All other lines represent the case of readout bias $\lambda_+ = 0.05$.

smaller negative peak followed by a blunt increase can be seen. In the case of the shortest stimulus (red curve), the second and third negative deflection partially overlap. Not surprisingly, the time course of $\hat{\mu}_{\mathrm{dr}}(t)$ is again mirrored in the case that $\lambda = \lambda_-$ (blue dashed line). The most prominent feature in the time course of $\hat{\sigma}_{\mathrm{dr}}(t)$ is a consistent increase simultaneous to the first peak in the mean activity (fig. 4.11E), when the bias is $\lambda = \lambda_+$. Importantly, the deviation of $\hat{\sigma}_{\mathrm{dr}}(t)$ stays slightly positive even during the sag of $\hat{\sigma}_{\mathrm{dr}}(t)$ at $t \approx 100\,\mathrm{ms}$. When $\lambda = \lambda_-$ the standard deviation is below the value of the spontaneous state in almost the entire duration of the stimulus (0 to 400 ms). The effect size measured by the DR detector (fig. 4.11F) is quite similar for all signals (when $\lambda = \lambda_+$) and in the range 2 % to 3 %. Although the mean readout activity, in the case that $\lambda = \lambda_-$, has the same number of favorable (one) and unfavorable peaks (two), the sustained decrease in $\hat{\sigma}_{\mathrm{dr}}(t)$ hampers the detectability, which results in a negative effect size.

Results from the DNR detector for $\hat{\mu}_{\mathrm{dnr}}(t)$ are qualitatively similar to those from the DR. However, the first peak and the first dip are smaller and deeper, respectively (fig. 4.11G, solid lines). The initial increase in $\hat{\sigma}_{\mathrm{dnr}}(t)$ is stronger, but lasts for a shorter time (fig. 4.11H, solid lines). The overall effect of these differences is that the effect size is rather small for all three signals (fig. 4.11I, solid lines). When the case of $\lambda = \lambda_-$ is considered, the first dip in $\hat{\mu}_{\mathrm{dnr}}(t)$ is reduced, while the first positive peak is larger than in the previous case (fig. 4.11G, dashed line), which should give a positive contribution to the effect size. However, the decrease in the standard deviation is much stronger (fig. 4.11H, dashed line), which results in a clearly negative effect size (fig. 4.11I, dashed line). Choosing $\lambda = \lambda_-$ leads to a negative effect size (from the DR and DNR detector) for all signals, including those discussed in the following sections (not shown), which is in contrast with the experimental results. Therefore, to shorten the presentation of results, this case will be discarded in the following.

In fig. 4.12, the effect size $\bar{\mathcal{Y}}_X$ obtained from simulation results is superimposed to the experimental data. Note that error bars are larger than in fig. 4.10 because of the substantially smaller size of the dataset. The effect size measured from the IR is generally larger than in the experiment and clearly grows as a function of the stimulation length (fig. 4.12A). The DR yields an effect that is independent of the stimulus duration. It is smaller in magnitude than in the data, but still compatible with the error bars (fig. 4.12B). The DNR measures an effect which is rather small for all stimuli.

The general picture for the case that $\mathcal{B}_0$ is a FS neuron presents some similarities with the case that a RS neuron is stimulated: the IR reliably detects the stimulation, but the effect strongly depends on the stimulus length, which is not the case in the experiments. The DR measures an effect that is essentially independent of the stimulus duration. However, the experimental data indicate that the effect size for FS neurons is generally larger than for RS neurons, a fact that is not observed in the model.

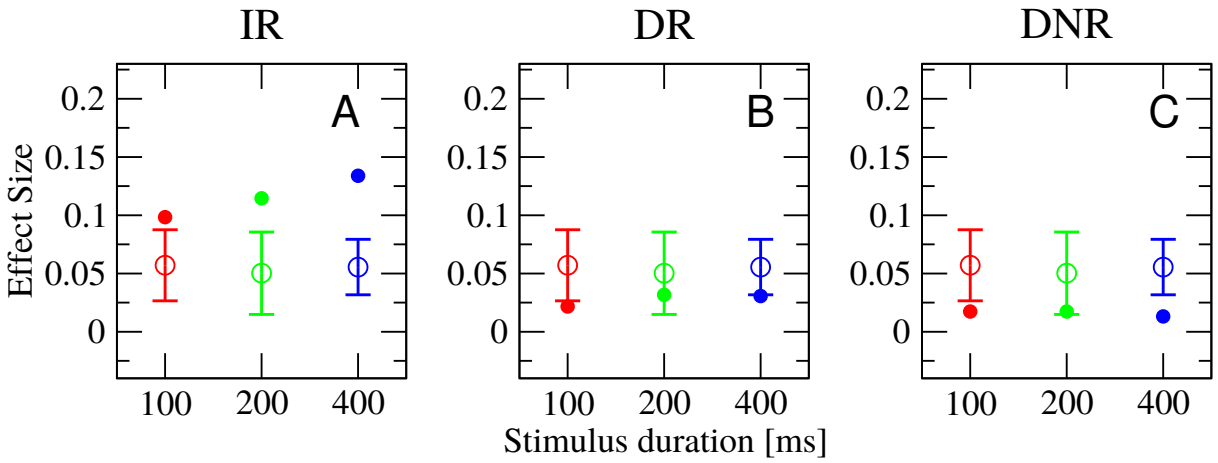

Figure 4.12. – **Results from differentiator readout (DR) and differentiator network readout (DNR) are generally lower than in the experimental data, whereas integrator readout (IR) yields results compatible in magnitude, but with inconsistent dependence on stimulus length (stimulation of FS cell, signals as in fig. 4.7A).** Open circles with error bars are experimental results, which are the same in each panel, and represent the average effect size computed from 18 FS cells. The number of trials per cell is rather heterogeneous (total number of trials is 394). Experimental data are from the Brecht lab. **A**, **B**, and **C**: filled dots are the effect size (for a false positive rate of 0.25) resulting from the IR, DR, and DNR, respectively.

4.2.2. Effect of stimulus intensity

In the second experiment, the firing rate of the stimulated cell was varied while keeping the number of elicited spikes roughly constant. To this end, the three stimuli shown in fig. 4.7B were used. The stimulus lasting 100 ms (red) has an intensity corresponding to 100% of the maximum current, the stimulus of length 200 ms (green) has intensity 50% of the maximum current, and the longest stimulus (400 ms, blue) has the lowest intensity (25% of the maximum current). In this way, the area (the amount of injected charge) stays constant. In the experiment, when the stimulated cell was a RS neuron, the three stimuli evoked a firing rate of (30 ± 10) Hz, (54 ± 23) Hz, and (109 ± 52) Hz, respectively. In the model, the evoked rates were higher on average, being (50 ± 12) Hz, (103 ± 20) Hz, and (150 ± 25) Hz. In the case that a FS cell was stimulated, the experimentally observed firing rates current injection were (60 ± 38) Hz, (120 ± 53) Hz, and (244 ± 100) Hz, for the 100, 200, and 400 ms stimulus, respectively. In the model, the same stimuli made the stimulated FS cell fire at (117 ± 18) Hz, (138 ± 21) Hz, and (153 ± 24) Hz, respectively. The maximum current in the model was chosen to achieve a good agreement in the evoked spike counts of the previous experiment. The general agreement of the average firing rates (which in the data is measured for a *different* set of cells) is less good than that of spike counts, but it is still mostly within one or two standard deviations from the experimental values.

The presentation of simulation results for the three stimuli of different intensity will proceed along the same lines as in the previous section. Starting with the case of current injection into a RS neuron, fig. 4.13 reports all detection statistics arranged in the same way as in the previous case. The same color code applies to the stimulus duration (red: 100 ms signal, green: 200 ms signal, blue: 400 ms signal, black: no signal). The only difference is that the stimuli of different length have different amplitude. Note that the 400 ms stimulus is identical to the previous case, so that simulation data were reused.

Figure 4.13A shows $\hat{\mu}_{\mathrm{ir}}(t)$ in response to stimuli of different frequencies. The initial small positive peak due to the spikes fired by $\mathcal{B}_0$ is similar for all stimuli, whereas the depth of the negative dip at ≈ 50 ms is larger for stronger stimuli, which is consistent with *in vitro* experiments of disynaptic (indirect) inhibition mediated by SOM-Martinotti interneurons (Silberberg and Markram, 2007). In these experimental findings, though, the latency of the inhibition also depends on the firing rate of the stimulated cell, which is not the case in fig. 4.13A. The width of the negative deflection is similar for all stimuli, after which $\hat{\mu}_{\mathrm{ir}}(t)$ relaxes to a plateau, except for the shortest stimulus, which only shows an inflection point before reaching the zero level. For all stimuli, $\hat{\mu}_{\mathrm{ir}}(t)$ overshoots when the stimulus is switched off before relaxing back to the spontaneous level. The qualitative course of $\hat{\sigma}_{\mathrm{ir}}(t)$ has a similar shape (fig. 4.13B), although the amplitude of the deviation from the spontaneous level is much smaller, especially in the case of the weaker signal (blue line). Interestingly, the effect size turns out to be similarly large for all three signals (fig. 4.13C). It is likely that the first strong negative deflection dominates the effect size for the two stronger signals, while the longer tail of the weakest stimulus compensates for the much smaller initial dip.

The mean DR activity shows a behavior which is qualitatively similar to the previous section (fig. 4.13D): after the small peak right at the stimulus onset, $\hat{\mu}_{\mathrm{dr}}(t)$ displays a downswing immediately followed by an upswing and a positive peak. A further positive peak appears when the stimulus is switched off, which partially overlaps with the previous one for the shortest signal (red curve). The main difference between the three signals and to the previous case is that both the first negative deflection and the following peak are increase in magnitude, when the stimulus grows in strength. The decrease in the standard deviation is rather large at the stimulus onset for the two stronger signals (fig. 4.13E, red and green) and the recovery to the spontaneous level is slower for the intermediate 200 ms stimulus. The effect size measured for the three stimuli is very similar (fig. 4.13F) and in the range observed in the data (2 % to 3 %).

The shape of $\hat{\mu}_{\mathrm{dnr}}(t)$, shown in fig. 4.13G, is qualitatively similar to that of $\hat{\mu}_{\mathrm{dr}}(t)$ for all three signals. However, two main differences can be noticed: first, the height of the first large positive peak around $t = 100$ ms is amplified, especially in the case of the two stronger signals; second, a broad dip can be seen after the stimulus is turned off, which corresponds to the final relaxation of the IR activity to the zero level and was barely present in the DR activity. The behavior

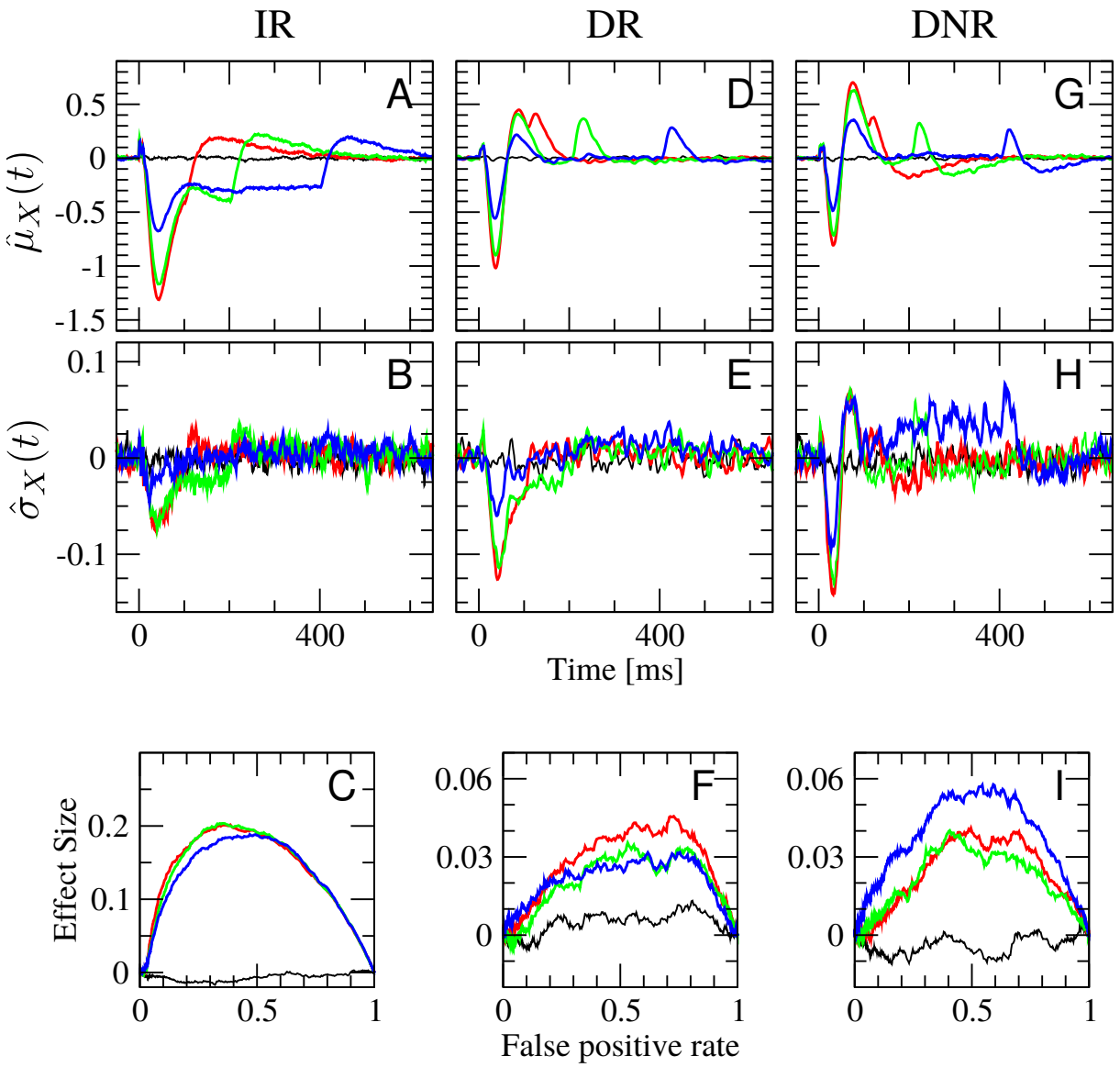

Figure 4.13. – **Summary of detection statistics upon stimulation of a RS cell with the three stimuli in fig. 4.7B (intensity inversely proportional to duration).** First row: standardized deviation from the spontaneous value of the time-dependent mean readout activity eq. (4.52). Second row: standardized deviation from the spontaneous value of the time-dependent standard deviation of the readout activity eq. (4.55). Third row: effect size as a function of the false positive rate. First column: integrator readout (IR). Second column: differentiator readout (DR). Third column: differentiator network readout (DNR). The color of each lines corresponds to a stimulus as in fig. 4.7A. Black line is catch trial condition (no stimulus).

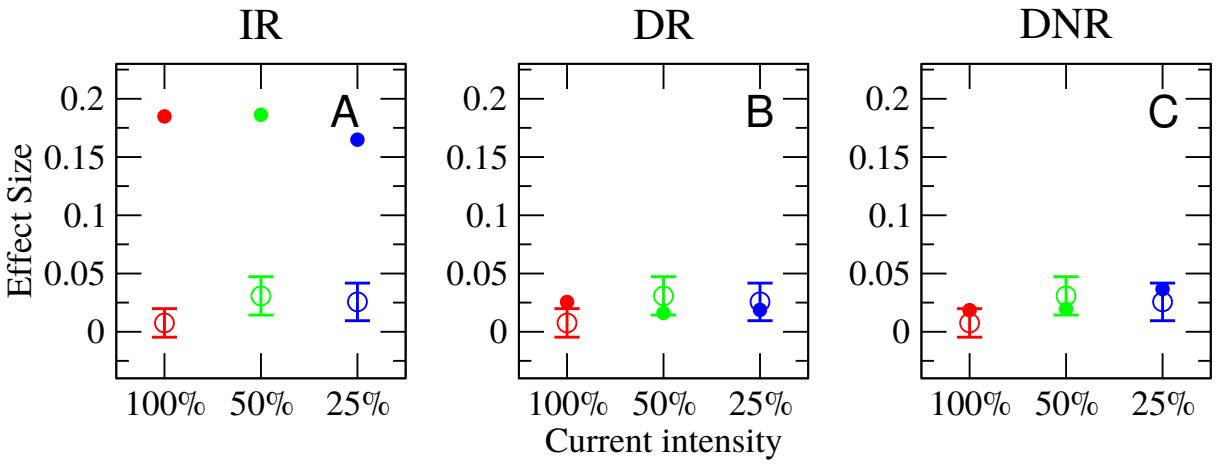

Figure 4.14. – **Results from all readouts show no pronounced dependence on the stimulus intensity, as the experimental data. However, the magnitude of the effect obtained from the differentiator readout (DR) and differentiator network readout (DNR) is similar to the experimental values, whereas the integrator readout (IR) yields a much larger effect (stimulation of RS cell, signals as in fig. 4.7B).** Open circles with error bars are experimental results, which are the same in each panel, and represent the average effect size computed from 55 RS cells. The number of trials per cell is rather heterogeneous (total number of trials is 1469). Experimental data are from the Brecht lab. **A**, **B**, and **C**: filled dots are the effect size (for a false positive rate of 0.25) resulting from the IR, DR, and DNR, respectively.

of $\hat{\sigma}_{\mathrm{dnr}}(t)$ (fig. 4.13H) shows positive and negative deflections from the spontaneous level in coincidence with those of the mean, which was not observed in the time-dependent standard deviation of the DR activity. All these differences compensate each other in the effect size for the signals of strong and intermediate strength, which is essentially identical to that measured for the DR (fig. 4.13I). As already noted above, the detectability for the 400 ms signal grows as a result of the sustained increase in $\hat{\sigma}_{\mathrm{dnr}}(t)$ for the weaker 400 ms signal.

Figure 4.14 directly compares the simulation results with the experimental data. All detectors yield an effect size which is only weakly dependent on the stimulus intensity. However, $\bar{\mathcal{Y}}_{\mathrm{ir}}$ (fig. 4.14A) is much larger than what observed in the data, while the magnitude of $\bar{\mathcal{Y}}_{\mathrm{dr}}$ (fig. 4.14B) and $\bar{\mathcal{Y}}_{\mathrm{dnr}}$ (fig. 4.14C) is compatible with that of the experimental measurements.

Simulation results for the case that a FS neuron is stimulated are shown in fig. 4.15. A comparison of fig. 4.15A to the equivalent plot of the previous section (fig. 4.11A) makes clear that differences are rather small. The duration of the three signals is the same, and the difference in the evoked firing rate is rather modest, as discussed at the beginning of this section. In fact, the weakest current intensity causes the neuron to fire at a rate which is already not far from saturation. Thus, when the stimulation current is further increased, the possible increase in firing rate is limited, which indicates that the value of the refractory period chosen for FS neurons was perhaps too large. This fact, combined with the strong depression of inhibitory

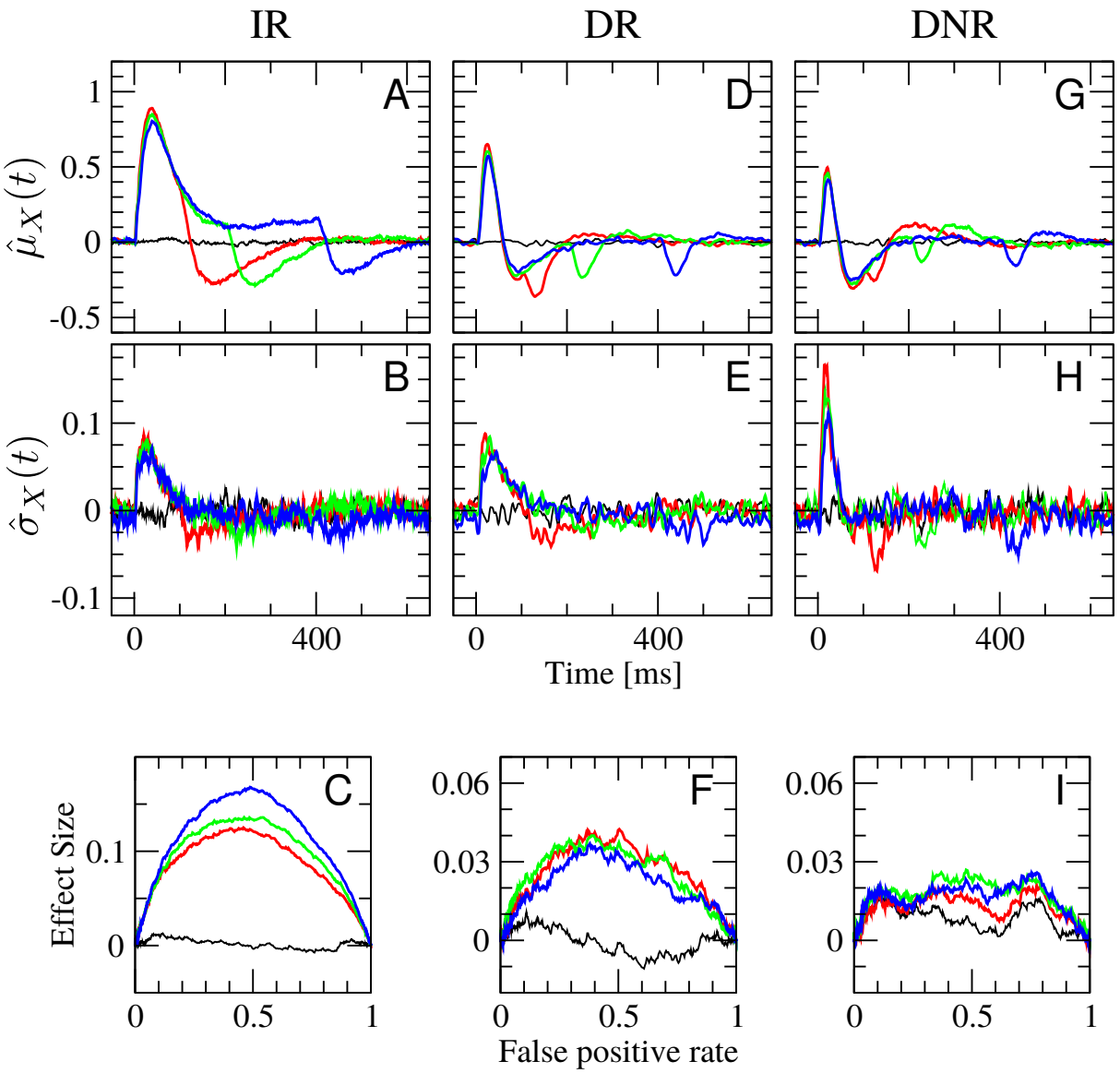

Figure 4.15. – **Summary of detection statistics upon stimulation of a FS cell with the three stimuli in fig. 4.7B (intensity inversely proportional to duration).** First row: standardized deviation from the spontaneous value of the time-dependent mean readout activity eq. (4.52). Second row: standardized deviation from the spontaneous value of the time-dependent standard deviation of the readout activity eq. (4.55). Third row: effect size as a function of the false positive rate. First column: integrator readout (IR). Second column: differentiator readout (DR). Third column: differentiator network readout (DNR). The color of each lines corresponds to a stimulus as in fig. 4.7A. Black line is catch trial condition (no stimulus).

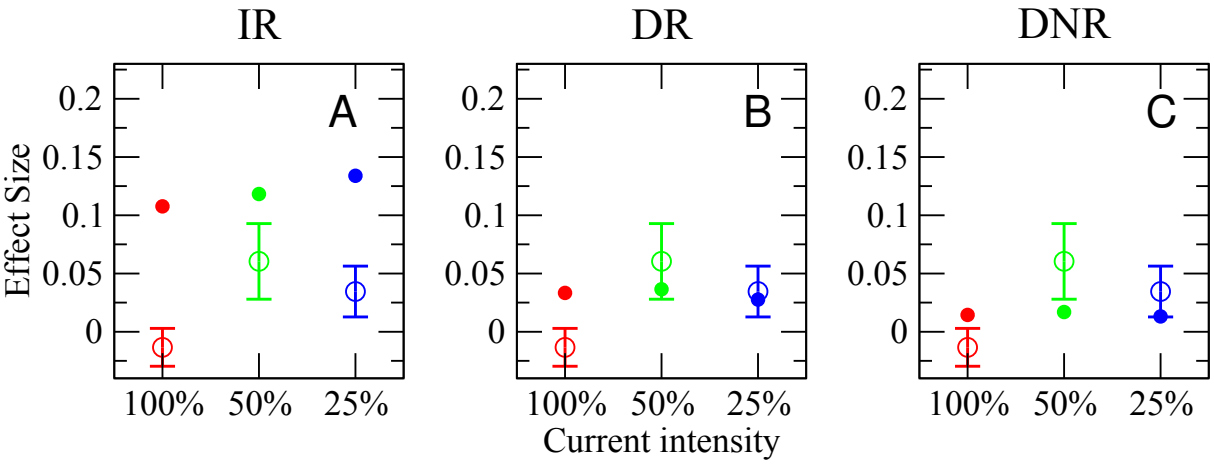

Figure 4.16. – **Results from all readouts show no clear dependence on the stimulus intensity, as the experimental data (note the larger error bars due to the limited size of this dataset). However, the magnitude of the effect obtained from the differentiator readout (DR) and differentiator network readout (DNR) is closer to the experimental values, whereas the integrator readout (IR) yields a much larger effect (stimulation of FS cell, signals as in fig. 4.7B).** Open circles with error bars are experimental results, which are the same in each panel, and represent the average effect size computed from 11 FS cells. The number of trials per cell is rather heterogeneous (total number of trials is 354). Experimental data are from the Brecht lab. **A**, **B**, and **C**: filled dots are the effect size (for a false positive rate of 0.25) resulting from the IR, DR, and DNR, respectively.

synapses of FS neurons causes the behavior of $\hat{\mu}_{\mathrm{ir}}(t)$ (fig. 4.11A) and of $\hat{\sigma}_{\mathrm{ir}}(t)$ (fig. 4.15B) for the three signals to be substantially the same as in the previously discussed case. Not surprisingly, the resulting effect size (fig. 4.15C) is also very similar.

Here, the initial peak and the subsequent dip in $\hat{\mu}_{\mathrm{dr}}(t)$ are slightly more pronounced for stronger signals (fig. 4.15D), and the same holds for $\hat{\sigma}_{\mathrm{dr}}(t)$ (fig. 4.15E), whose initial maximum is a little stronger and earlier when the stimulus intensity is larger. These (in this case rather limited) changes compensate each other when the effect size of the three signals is considered, in which no appreciable difference between the signals is seen (fig. 4.15F).

A similar picture is observed for the DNR activity. Deviations from the zero level in both directions are slightly larger for stronger signals in the time course both of $\hat{\mu}_{\mathrm{dnr}}(t)$ (fig. 4.15G) and of $\hat{\sigma}_{\mathrm{dnr}}(t)$ (fig. 4.15H). However, they do not produce any appreciable influence on the effect size, which is very close to the noise floor for all three signals (fig. 4.15I).

The experimental dataset for this condition is rather small (see caption to fig. 4.16), so that it is possible that the non-monotonic dependence on the stimulus intensity is due to the large measurement noise, and that the measured effect sizes are not precise. Still, the effect size measured by the IR can be likely regarded as too large (fig. 4.16A), as in the previous cases. The DR yields an effect which is within the error bars in two out of three cases (fig. 4.16B) and

the DNR (fig. 4.16C) underestimates the effect size.

4.2.3. Effect of stimulus regularity

In the third and final experiment modeled here, random stimuli were employed to elicit irregular spike trains. These stimuli consisted of a random sequence of six current steps of length 10, 20, 40, 80, 160, and 90 ms, with current intensity 100%, 50%, 25%, 12.5%, 6.25%, and -50%, respectively. In other words, each sequence always consists of five positive (depolarizing) current steps with intensity inversely proportional to the duration and of one negative step of 90 ms, which inhibits the cell from firing. These six steps were randomly shuffled in each trial. Two example signals are shown in fig. 4.7C. The total duration of each irregular stimulus was of 400 ms, and the detectability of these stimuli was compared to that of regular steps of 400 ms at 25% of the maximum current, used in the previous conditions. In the experiment, irregular current injections into RS neurons generated spike trains with an average firing rate of 24 Hz and CV ≈ 1.1. In the model, the average rate was 28 Hz and the average CV ≈ 1.3.

Figure 4.17 shows the overview of all results obtained from the three detectors. The blue line refers, as in all previous plots, to the 400 ms regular stimulus. The response to the 400 ms stimulus has already been described twice, because this signal appears in both sets of stimuli used in the previous experiments. Therefore, blue lines are provided as reference for the comparison with the response to irregular stimuli but will be not discussed in detail as such. Two sets of $N_{\text{trials}} = 10000$ random realizations of the $6! = 720$ possible stimuli were used. Results for the first set of stimuli are plotted with red lines. Orange lines indicate results for the second set of irregular stimuli. Comparing red and orange lines is useful to estimate the variability due to the random choice of signals.

The initial time course of $\hat{\mu}_{\text{ir}}(t)$ in response to irregular stimulation is similar in shape to that of the regular one, but both the first dip and the value of the plateau are smaller in magnitude (fig. 4.17A), which is not surprising as the total amount of injected current (and of elicited spikes) is also smaller. However, at the end of the stimulation, $\hat{\mu}_{\text{ir}}(t)$ displays a sharp decrease before relaxing back after the stimulus is switched off. This negative dip corresponds to a peak in the average firing rate of the SOM population, which is caused by trials in which the negative step appears in the middle and a strong positive step occurs at the end. In this case, a combination of the different time scales of synaptic facilitation, depression, and spike-frequency adaptation causes a response which is stronger than the initial one. This strong response occurs when $\mathcal{B}_0$ has already fired some spikes during a first positive current step, which increases the strength of the facilitating synapses but also causes a buildup in the adaptation current of $\mathcal{B}_0$ and of SOM neurons. When the negative current injection silences the cell, the spike-frequency adaptation variables recover much faster than the synaptic facilitation because of the different time scales. Therefore, when the cell resumes firing, the inhibitory effect of the adaptation has

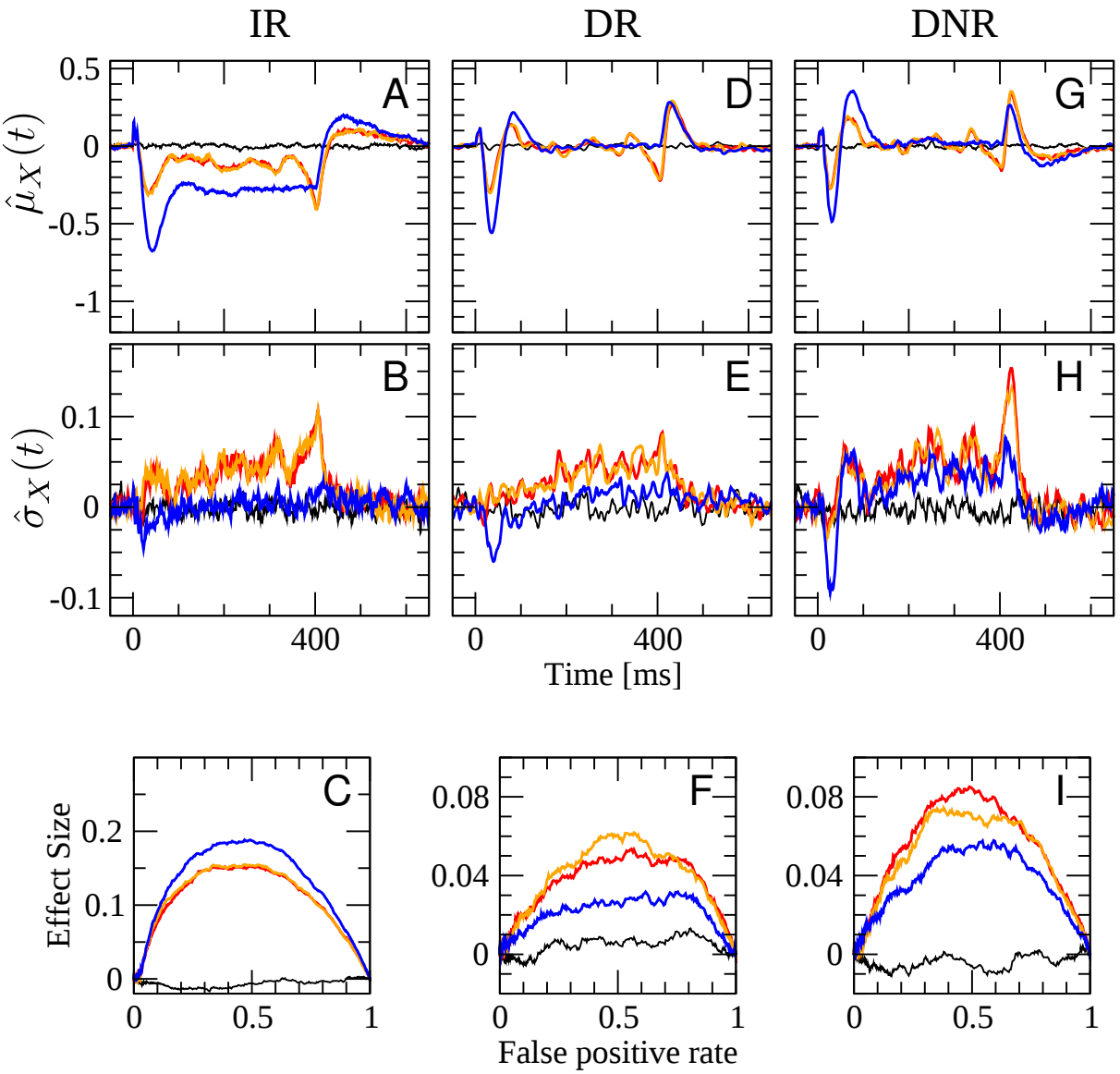

Figure 4.17. – **Summary of detection statistics upon stimulation of a RS cell with irregular three stimuli as in fig. 4.7C (random permutation of multiple current steps), compared to regular** 400 ms **stimulation.** First row: standardized deviation from the spontaneous value of the time-dependent mean readout activity eq. (4.52). Second row: standardized deviation from the spontaneous value of the time-dependent standard deviation of the readout activity eq. (4.55). Third row: effect size as a function of the false positive rate. First column: integrator readout (IR). Second column: differentiator readout (DR). Third column: differentiator network readout (DNR). Blue line refers to regular stimulus, as in the previous cases. Red and orange lines represent two different random samples of 10000 stimuli (stimulus is changed in each trial) from the 720 possible permutations of the six steps. Black line is catch trial condition (no stimulus).

vanished, while the amplitude of the facilitation variables is still above the spontaneous value, thus causing a strong response.

It may seem strange to consider the average over different irregular stimuli that produce very heterogeneous responses. In fact, the mean response conditional on each of the frozen irregular stimuli alternates peaks and troughs and looks very different from the response seen in fig. 4.17A. In particular, positive and negative current steps occurring at different positions within the 400 ms partially compensate each other in the trial average, which is the main cause for the rather weak mean response in the middle of the stimulation time window. However, rats in the actual experiments do not know which particular realization of the irregular sequence was being used in each trial. Therefore, the ensemble of trajectories used here, in which the stimulus is changed in each trial, correctly represents the experimental situation and it is legitimate to consider its time-dependent mean. Still, the picture offered by the time-dependent mean alone would be highly misleading in this case, because the variability in the input signal, which averages out in the mean, manifests itself in an increase in the time-dependent standard deviation, as it can be seen in fig. 4.17B. Indeed, $\hat{\sigma}_{\mathrm{ir}}(t)$ grows steadily over the entire time interval in which the stimulus is active and peaks at the end. The increase in variance facilitates the detectability of the irregular stimulus, as shown by fig. 4.17C: despite the much smaller deviation from the mean, the effect size measured for irregular stimuli is almost as large as that of the regular signal.

As in the previous cases, the DR is zero during the plateau of $\hat{\mu}_{\mathrm{ir}}(t)$, and $\hat{\mu}_{\mathrm{dr}}(t)$ shows up- and downswings at the same place as the regular signal (fig. 4.17D). The amplitude of the first dip and of the first peak is smaller than for the regular signals, and an additional negative peak is present at $t \approx 400\,\mathrm{ms}$, which is likely to negatively affect the detectability of irregular signals. However, the comparison of fig. 4.17E with fig. 4.17B reveals that the increase in standard deviation is almost unaffected by the DR. As a result, $\hat{\sigma}_{\mathrm{dr}}(t)$ increases steadily over the entire duration of the stimulation, which allows the irregular stimulus to achieve an effect size that is twice as large as that of the regular stimulus for most values of the false positive rate (fig. 4.17F).

Figure 4.17G shows that $\hat{\mu}_{\mathrm{dnr}}(t)$ is very similar to $\hat{\mu}_{\mathrm{dr}}(t)$, as opposed to $\hat{\sigma}_{\mathrm{dnr}}(t)$, which displays some differences with respect to the DR (fig. 4.17H). As already noticed in the previous cases, the variability of the DNR activity is larger than that of the corresponding DR activity, which results in an increased effect size for the regular stimulus. However, $\hat{\sigma}_{\mathrm{dnr}}(t)$ for the irregular stimulus is always above that of the regular signal, and the final peak is particularly strong. Consequently, the effect size for both signal types is increased in comparison to the DR (fig. 4.17I). Although the relative difference between the curves is smaller, the irregular stimulus is still clearly more effective.

Figure 4.18 shows that the difference in the effect size of regular and irregular signals observed experimentally is rather large. As in the previous cases, results obtained from the IR are in

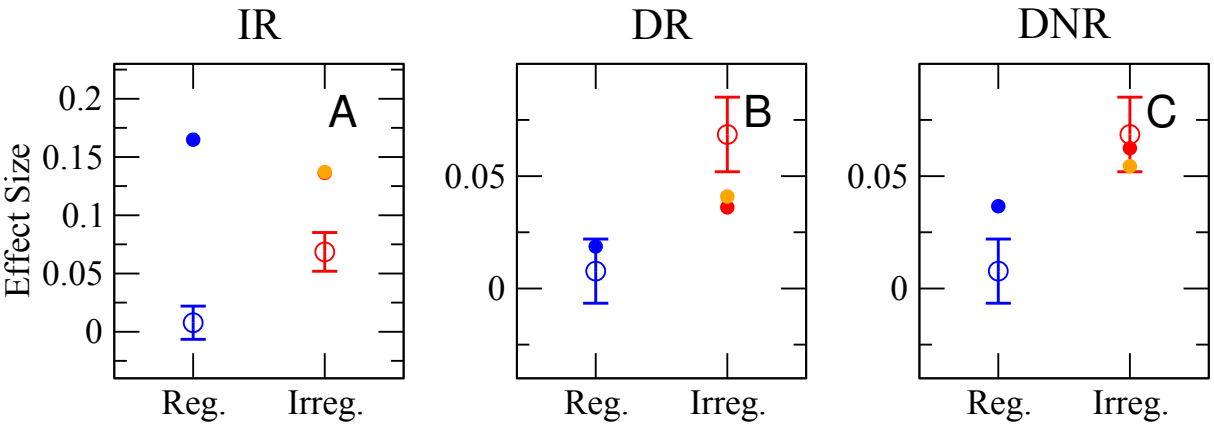

Figure 4.18. – **Effect size measured from differentiator readout (DR) and differentiator network readout (DNR) is larger for irregular stimuli than for regular stimuli, which is consistent with the experimental data, whereas the opposite holds for the integrator readout (IR) (stimulation of RS cell, irregular signals are as in fig. 4.7C.** Open circles with error bars are experimental results, which are the same in each panel, and represent the average effect size computed from 62 RS cells. The number of trials per cell is rather heterogeneous (total number of trials is 1780). Experimental data are from the Brecht lab. **A**, **B**, and **C**: filled dots are the effect size (for a false positive rate of 0.25) resulting from the IR, DR, and DNR, respectively. Red and orange circles are two different random samples of irregular signals.

disagreement with the data, as the effect size in response to irregular stimuli is smaller than that relative to regular ones (fig. 4.18A). On the contrary, both the DR and the DNR register a clear increase in effect size if irregular stimuli are used in place of regular ones (fig. 4.18B and C, respectively). In both cases, however, the difference is smaller than in the data. The DR underestimates the effect of irregular stimuli, while the DNR overestimates the effect of regular ones.

Finally, the effects of stimulating a FS neuron with irregular stimuli will be considered. As shown in fig. 4.19A, the initial response of the time-dependent mean of the IR activity is, on average, weaker than that to the regular stimulus, but it saturates around a similar value before undershooting at $t \approx 400\,\mathrm{ms}$ and finally relaxing to the spontaneous value. As discussed before, the stimulus-related variability manifests itself in the marked increase in $\hat{\sigma}_{\mathrm{ir}}(t)$ (fig. 4.19B), which facilitates the detection. However, it does not suffice to make the irregular stimulus more easily detectable than the regular one, as seen in fig. 4.19C.

The time course of $\hat{\mu}_{\mathrm{dr}}(t)$ and of $\hat{\mu}_{\mathrm{dnr}}(t)$ in response to an irregular stimulus are very similar to each other and to the response to the regular stimulus (fig. 4.19D and G). However, the DR and the DNR react differently to the stimulus-induced variability: the increase in $\hat{\sigma}_{\mathrm{dr}}(t)$ displays an initial peak and a long tail (fig. 4.19E), similarly to the IR activity, whereas $\hat{\sigma}_{\mathrm{dnr}}(t)$ rapidly decreases to the baseline level after the sharp peak at the stimulus onset (fig. 4.19H). This difference between the two detectors is reflected in the effect size, which is larger for the DR

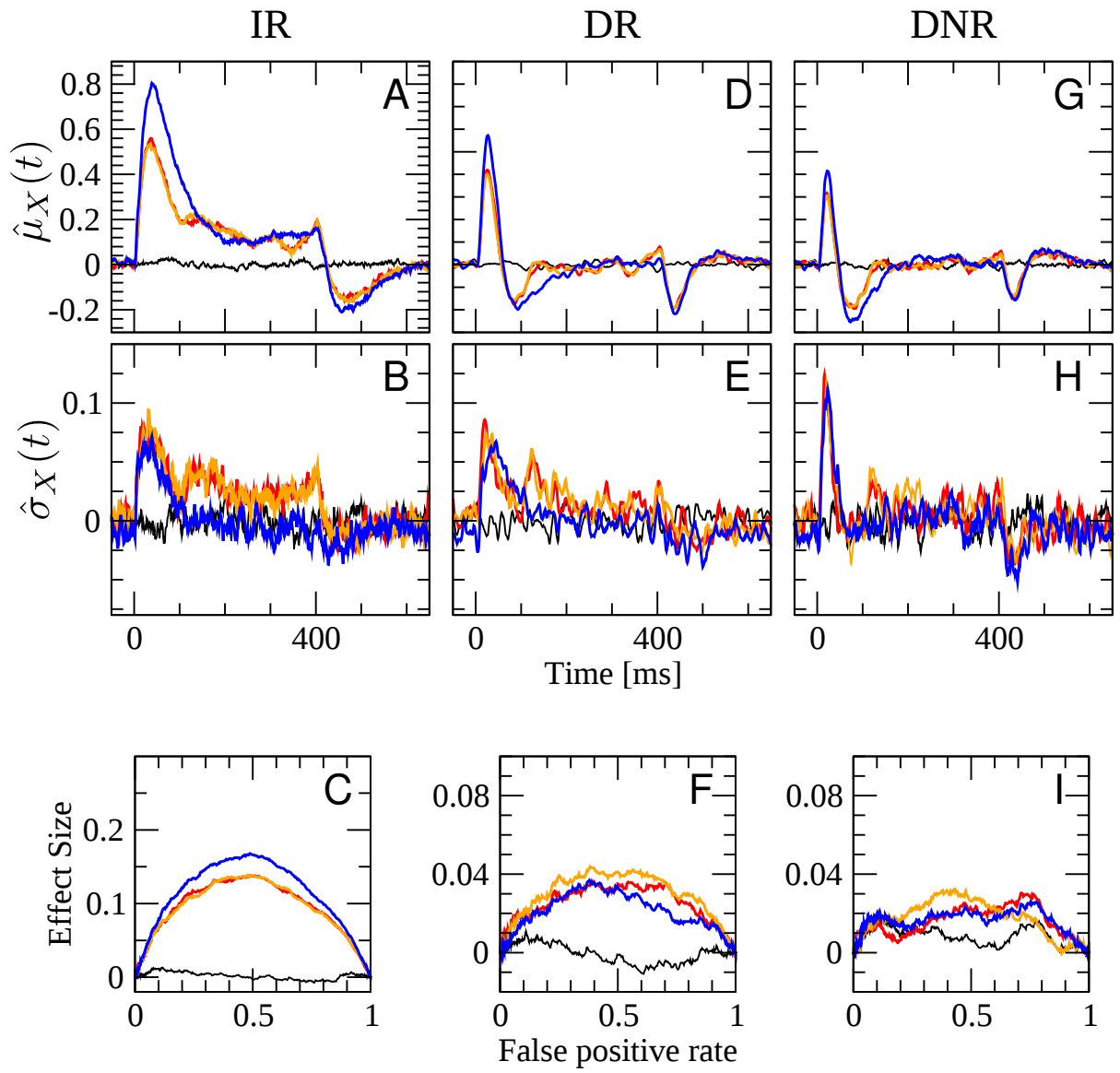

Figure 4.19. – **Summary of detection statistics upon stimulation of a FS cell with irregular three stimuli as in fig. 4.7C (random permutation of multiple current steps), compared to regular** 400 ms **stimulation.** First row: standardized deviation from the spontaneous value of the time-dependent mean readout activity eq. (4.52). Second row: standardized deviation from the spontaneous value of the time-dependent standard deviation of the readout activity eq. (4.55). Third row: effect size as a function of the false positive rate. First column: integrator readout (IR). Second column: differentiator readout (DR). Third column: differentiator network readout (DNR). Blue line refers to regular stimulus, as in the previous cases. Red and orange lines represent two different random samples of 10000 stimuli (stimulus is changed in each trial) from the 720 possible permutations of the six steps. Black line is catch trial condition (no stimulus).

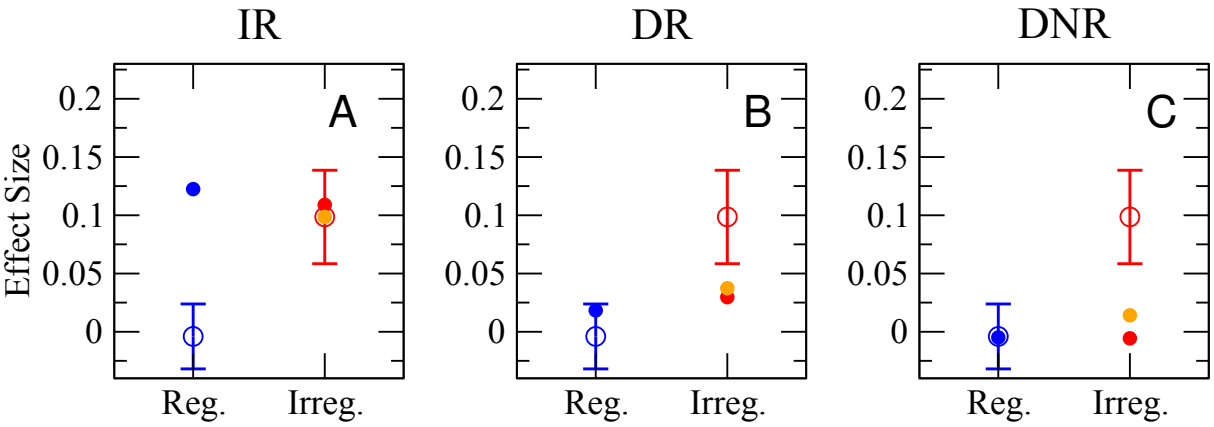

Figure 4.20. – **Effect size measured from differentiator readout (DR) and differentiator network readout (DNR) is slightly larger for irregular stimuli than for regular stimuli, but the difference is smaller in magnitude than in the experimental data, whereas the integrator readout (IR) yields an effect that is equally large for regular and irregular stimulation, unlike the experimental results (stimulation of FS cell, irregular signals are as in fig. 4.7C).** Open circles with error bars are experimental results, which are the same in each panel, and represent the average effect size computed from 12 FS cells. The number of trials per cell is rather heterogeneous (total number of trials is 389). Experimental data are from the Brecht lab. **A**, **B**, and **C**: filled dots are the effect size (for a false positive rate of 0.25) resulting from the IR, DR, and DNR, respectively. Red and orange circles are two different random samples of irregular signals.

(fig. 4.17F) than for the DNR (fig. 4.17I). Although only slightly, the effect caused by irregular stimuli exceeds that generated by regular ones.

Figure 4.20 compares the experimental data to the effect size measured by the three readout models. The effect size measured experimentally undergoes a large increment when irregular stimulation is used instead of the regular one. However, the measurement noise is here rather large because of the small dataset (note that the effect size for regular stimuli is here zero, while it was about 5% in other datasets). The IR is in disagreement with the data, because it detects more reliably the regular stimulation than the irregular one (fig. 4.20A). The effect of the regular stimulus measured by the DR is compatible with the experimental data. Furthermore, the DR detects irregular stimuli slightly better (fig. 4.20B), which is in qualitative agreement with the experiment, although the increase seen in the data is much larger. The effect resulting from the DNR is very weak and the faint increase for irregular stimuli is observed only in one of the two sets of stimuli and it is therefore probably due to a fluctuation.

4.3. Summary and discussion

One way of summarizing the experimental results by Doron (2012); Doron et al. (2014) is that the response rate of the rat is not appreciably influenced by the strength of the stimulation, interpreted both in terms of number and frequency of the elicited spikes, whereas a variable stimulation evoking an irregular spike train has a significantly stronger impact than a constant stimulus. The purpose of this chapter was to explore what attributes of the network and readout model can render the simulation results compatible with these experimental findings.

The first strategy to achieve this goal was to equip the recurrent network model with more details about the actual properties of neurons and of the synaptic transmission in the barrel cortex (fig. 4.3, p. 139). In the previous chapters, excitatory and inhibitory cells had the same properties, except for the effect on their targets. However, excitatory neurons and the most common type of inhibitory neurons, i.e. fast-spiking (FS) interneurons, are different in their cellular properties and connection patterns. One basic difference is that FS neurons have faster time constants and higher spontaneous firing rates than excitatory regular-spiking (RS) cells. Furthermore, they form more dense connections with other neurons, but on a shorter range. To construct a model approximately consistent with these facts without introducing a space-dependent connection probability, which would radically change the network's dynamics (Rosenbaum et al., 2017), the network was considerably downscaled to a size representing only the immediate vicinity of the stimulated cell (within a radius of $\approx 200\,\mu m$).

Although FS neurons may be seen as the backbone of the cortical inhibitory system in terms of the larger number of neurons and fired action potentials, many more classes of inhibitory neurons have been identified and classified in the cortex (Tremblay et al., 2016). Among these inhibitory cell types, the somatostatin-expressing low-threshold spiking (SOM-LTS) neurons stand out both for their relative abundance in the barrel cortex (Meyer et al., 2011) and for their functional properties (Beierlein et al., 2003). In particular, the combination of strong facilitating input synapses and the low firing-threshold makes them likely to play an important part in the network response when a pyramidal excitatory cell is forced to fire in bursts, as several experimental studies suggest (Silberberg and Markram, 2007; Kapfer et al., 2007; Kwan and Dan, 2012). Therefore, this cell class was included in the new network model. Because of the central role of short-term synaptic plasticity in the function of these neurons, it was necessary to incorporate synaptic dynamics into the model.

Although short-term facilitation is not an exclusive property of connections from excitatory RS cells to SOM-LTS neurons, the majority of synapses in the barrel cortex display pronounced short-term depression (STD) (Beierlein et al., 2003; Cowan and Stricker, 2004; Feldmeyer et al., 2005; Helmstaedter et al., 2008; Lefort and Petersen, 2017). Therefore, all other synapses in the network were modeled as decisively depressing, a mechanism that could be expected to

naturally suppress the effects of a constant stimulus. The same expectation also applies to the strong spike-frequency adaptation that both excitatory RS and SOM-LTS neurons exhibit (Gottlieb and Keller, 1997; Beierlein et al., 2003), and that was also included in the model.

A final addition to the model is represented by the presence of gap junctions between inhibitory cells of the same kind (Galarreta and Hestrin, 1999; Beierlein et al., 2000; Amitai et al., 2002). Gap junctions are believed to promote synchrony and oscillations in the network dynamics (Beierlein et al., 2000) and are likely to be involved in the long-lasting inhibitory effect of cortical microstimulation (Butovas et al., 2006). Here, gap junctions were mimicked by a fast global excitatory spiking coupling. Gap junctions between SOM neurons can amplify the response to excitatory input from the stimulated RS cell, while gap junctions between inhibitory FS cells could synchronize and amplify the response of the FS population to the direct stimulation of a FS neuron. However, a strong excitatory coupling between inhibitory cells also promotes the formation of noisy oscillations, which reduce the signal-to-noise ratio. If the coupling exceeds a critical value, a complete synchronization of the inhibitory neurons occurs (not shown).

The second strategy employed to seek a congruence between the model results and the experiments was to modify the readout procedure. The readout schemes introduced in the last chapters acted as integrators of the firing activity of a subset of the network. The first of the three readout schemes considered in this chapter, the integrator readout (IR), operates similarly (fig. 4.2A, p. 137). The main difference from the readout of the previous chapters is that the spikes of the readout population (a subset of the recurrent network) are not filtered with a static synaptic profile, but obey the same STD equation as recurrent network connections, thus introducing a mechanism that can potentially oppose changes in the input. However, results obtained from the IR disagree with the experimental observations in several respects: in all considered cases, the effect of single-cell stimulation measured by the IR was large (mostly above 10%, and in several cases close to 20%) and depended strongly on the duration and intensity of the constant stimulus. Furthermore, irregular stimuli were less detectable than their regular counterparts. Hence, none of the model ingredients described above seems to be sufficient to explain either the insensitivity of the actual readout to the strength of the constant stimulation or the preference for irregular stimuli, when the IR readout is employed.

The second readout scheme used in this chapter reacted to differences between the activity of the readout population evaluated at two time points (fig. 4.2B), which accounts for its designation as the differentiator readout (DR). With this readout scheme, most aspects of the experimental data could be reproduced in the case that an excitatory RS neuron is stimulated. When a constant current injection was employed, the effect size was in the range 2 % to 3 % and it was essentially independent of the stimulation length and intensity, which is in good agreement with the data. Furthermore, the irregular current injection doubled the response rate with respect to the regular one, which is in qualitative agreement with the experimental observations,

in which the increase is, albeit, even larger. In the case that the single-cell stimulation is applied to a FS neuron, results still agree with the data in that the effect size does not depend on the length of intensity of the current. However, in contrast to the experimental findings, there is also no substantial difference between the effects of regular and irregular stimulation. Furthermore, the effect size tends to be somewhat smaller than in the data.

The third readout scheme (fig. 4.2C) was termed the differentiator network readout (DNR) because it is an attempt at implementing the DR with an integrate-and-fire network. Its basic architecture is equivalent to the readout network considered in section 3.3: it consists of one excitatory RS population ($\mathcal{S}^B$) and of one inhibitory FS population ($\mathcal{I}$), both receiving the same input from the RS neuron in the network in which the single-cell stimulation is applied. The inhibitory readout population $\mathcal{I}$ provides both recurrent inhibition to itself and feed-forward inhibition to $\mathcal{S}^B$. As explained in section 4.1.4, the strength of the feed-forward inhibition was tuned to balance as precisely as possible the direct feed-forward excitation. Crucially, the longer transmission delays of the inhibitory path cause the cancellation to occur at a later time, so that the DNR approximately realizes the same operation of the DR. Indeed, the average DNR activity was rather similar to the average DR in all cases. As a consequence, the response rates of the DNR were qualitatively analogous to those of the DR. In some cases, however, the time-dependent standard deviation of the DNR activity in response to the simulation was different from that of the DR, which caused some minor quantitative discrepancies in the effect size measured by the two readouts. It is not straightforward to precisely identify the source for the different behavior of the time-dependent standard deviation, but it may be due to a non-linear amplification of input cross-correlations. Perhaps, a larger amount of external white noise in the DNR might linearize its response and render it more similar to that of the DR. However, despite the minor differences, the DNR yielded results similar to those of the DR, which correctly captured several features of the dataset, especially in the case of the stimulation of a RS cell.

As discussed in section 4.1.4, a particular choice of the synaptic coupling parameters and delays are needed to make the DNR operate as a differentiator circuit. It is natural to assume that this parameter configuration can result from the training phase. It is also possible that a suitable learning algorithm can optimize parameters better than the simple criterion used here, thus better approximating the differentiation operation. Concerning the length of the time lag ΔT used to "differentiate" the network activity, it must be noted that the instantaneous current jumps of the signal cause changes in firing rate of the RS population that, for a network in the asynchronous state, occur on time scales on the order of the membrane time scale of the single neurons (see the discussion in section 2.2 and the reference to Gerstner et al., 2014). Hence, to obtain a good signal-to-noise ratio, the time lag should be set to a value on the same order of magnitude of the RS neuron membrane time constant. Is the value chosen here

$\Delta T = 10\,\text{ms}$ biologically plausible? It has been experimentally determined that the latency between a presynaptic spike and the PSP onset in the rat barrel cortex is approximately a linear function of the distance with slope 5 ms/mm (Helmstaedter et al., 2008). Therefore, a lag of 10 ms corresponds to an inter-somatic distance of $\approx 2\,\text{mm}$, which is a large, but not unphysiological value (Schnepel et al., 2014). Furthermore, although in the model the additional time lag of the inhibitory path was entirely assigned to the connections from the BCN to the inhibitory readout $\mathcal{I}$, dividing the latency evenly among the connections from the BCN to $\mathcal{I}$ and those from $\mathcal{I}$ to $\mathcal{S}^B$ would produce no change at all in the network's dynamics and ultimately yield the same results. In this way, the inhibitory signal would, for instance, need to travel back and forth from a population only one millimeter away.

In chapters 2 and 3, the detectability of inhibitory and excitatory cells was different in magnitude, but was based on the same mechanism and set the same requirement to achieve a given effect size, i.e. a readout biased towards cells receiving direct input from the stimulated cell. This symmetry is not surprising, because in the network models of the previous chapters excitatory and inhibitory neurons were identical except for the strength of their outgoing synapses, which explained the larger effect size measured when stimulating an inhibitory neuron. In the model of this chapter, the numerous differences in cellular parameters and connectivity between excitatory RS neurons and inhibitory FS neurons led to substantial differences in the response of the network in the two cases.

When a RS neuron is stimulated, only a modest fraction of the surrounding excitatory neurons receives direct excitatory input, which is also dampened by the STD of RS-to-RS connections, whereas a large fraction of the SOM-LTS neurons is activated owing to the strong facilitation of the RS-to-SOM synapses (fig. 4.5A, p. 151). As a consequence, stimulating an excitatory cell has mainly an inhibitory effect on the surroundings, which is due to the activation of the SOM-LTS neurons. This apparently contradictory conclusion is in accordance with experimental evidence (Silberberg and Markram, 2007; Kapfer et al., 2007; Kwan and Dan, 2012) and proved to be an effective way of generating a detectable population signal without the need for a readout bias.

When a FS neuron is stimulated, a large fraction of the local RS population receives direct inhibition and reduces its average firing rate. However, the stimulated neuron also inhibits a large fraction of the FS inhibitory population, which, in turn, provides less input to the RS population (fig. 4.5B). Consequently, RS neurons that do not receive direct inhibition increase their firing rate, thus compensating the decrease in the rest of the RS population. This circumstance makes a strongly biased readout necessary to achieve a modest effect size, which may cast some doubt on the plausibility of the model, as far as the detectability of FS neurons is concerned. Although experimental evidence reports dense and strong chemical inhibitory synapses connecting FS neurons with each other (Gibson et al., 1999; Beierlein et al., 2003; Avermann et al., 2012; Pfeffer et al., 2013), it would be a tempting idea to reduce the strength of the FS-to-FS coupling to

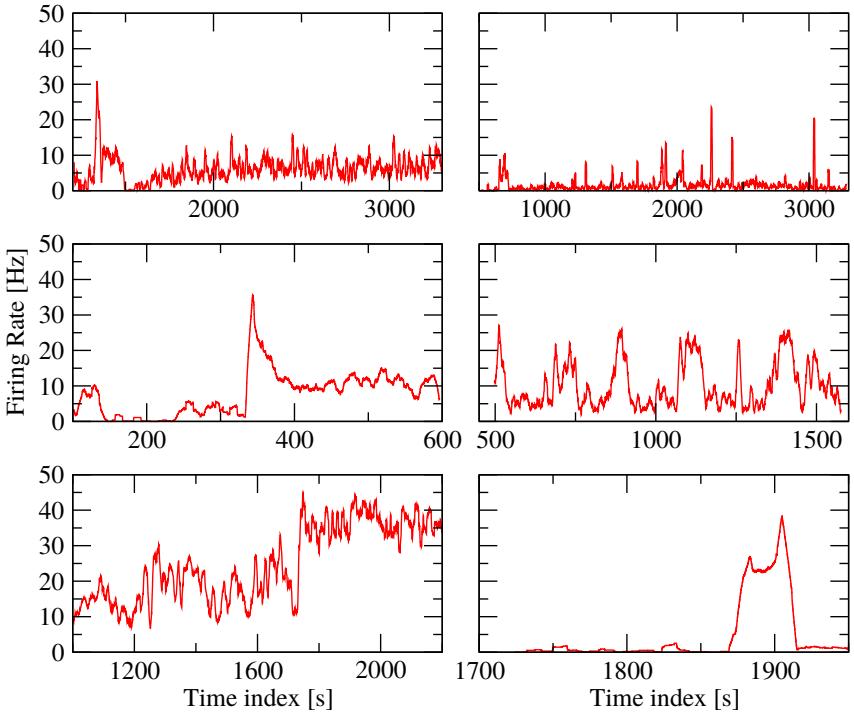

Figure 4.21. – Non-stationarity in the firing rate of cells measured *in vivo* in the barrel cortex. Firing rate (running average with 1 s time window) measured from selected neurons during the single-cell simulation sessions. Data are from the Brecht lab.

prevent the disinhibition that compensates the inhibition caused by the stimulated FS neuron. However, weakening FS-to-FS connections destabilizes the asynchronous irregular state, thus increasing the readout noise and limiting detectability. It is possible that an optimal point within a physiologically plausible parameter range exists, that is, where the sensitivity to the stimulation of a single FS neuron is increased without compromising the stability of the asynchronous state. Unfortunately, an extensive parameter search is impractical because of the large dimensionality of the parameter space and the large number of network simulations necessary for each point in the parameter space. Another possibility is that the better detectability of FS cells observed in the experiments relies on biological mechanisms absent in the model. One candidate mechanism may be, for instance, the activation of slow inhibitory receptors ($GABA_B$), which are involved in the late inhibitory response to microstimulation (Butovas et al., 2006).

Taken together, the results of this chapter indicate that a readout circuit operating as a differentiator is in better agreement with the experimental data than an integrator readout. However, the integrator readout detected the single-cell stimulation more reliably in all situations, a fact which has a puzzling implication. Recalling that the training protocol rewards correct detections, why should animals develop a sub-optimal readout and miss valuable saccharine water drops? One possible answer is that although an integrator readout may be more reliable in

the perfectly stationary state considered here, it would not robustly operate in the highly non-stationary situation that the actual readout must face. Figure 4.21 shows the spontaneous firing rate of some neurons during single-cell stimulation recording sessions in the barrel cortex (data from the Brecht lab). The strong and diverse variations in firing rate are mostly unrelated to any stimulation event, but are due to variations in the network state. A readout integrating spikes in a sliding time window would encounter severe difficulties in distinguishing stimulation events from the background variability, whereas a differentiator readout might effectively separate slow fluctuations from the faster changes induced by the stimulation.

Another potential advantage of a differentiator readout over an integrator may be related to the fact that the training of the readout occurs by using microstimulation, which induces additional spikes in the local network initially, but strongly suppresses firing shortly thereafter (Butovas et al., 2006; Histed et al., 2009). Hence, it is possible that searching for abrupt transitions from firing to silence and back is a more reliable way to detect such effects than to integrate over a longer window, in which the initial increased firing and the later inhibitory response can compensate each other. In this interpretation, the irregular single-cell nanostimulation generates a spike train that resembles more closely the effects of microstimulation, the signal with which the detector was trained.

4.4. Table of parameters

The following tables list all parameters used in this chapter with their numerical value. Angular brackets mark parameters that are drawn independently from a random distribution. In this case, the second column reports the mean. The standard deviation is indicated by σ, when necessary (it is equal to the mean for the exponential distribution). A different convention is used for the uniform distribution, of which minimum (maximum) value are reported. The third column provides a short description of the symbol and the type of probability distribution, when applicable.

Single-neuron parameters

Symbol	Value	Description
$\langle \tau_{m,E} \rangle$	20 ms, $\sigma = 20\%$	membrane time constant (RS), lognormal
$\langle \tau_{m,S} \rangle$	20 ms, $\sigma = 20\%$	membrane time constant (SOM-LTS), lognormal
$\langle \tau_{m,I} \rangle$	10 ms, $\sigma = 20\%$	membrane time constant (FS), lognormal
$\tau_{\mathrm{ref},0}$	4 ms	refractory period (constant part)
$\langle \hat{\tau}_{\mathrm{ref}} \rangle$	2 ms, $\sigma = 1$ ms	refractory period (variable part, lognormal)
$\langle v_{T,E} \rangle$	20 mV, $\sigma = 2$ mV	threshold voltage (RS), Gaussian
$\langle v_{T,I} \rangle$	20 mV, $\sigma = 2$ mV	threshold voltage (FS), Gaussian
$\langle v_{T,S} \rangle$	14 mV, $\sigma = 1$ mV	threshold voltage (SOM-LTS), Gaussian
v_R	10 mV	reset voltage
$\langle \tau_{a,E} \rangle$	100 ms, $\sigma = 20\%$	adaptation time constant (RS), lognormal
$\langle \Delta a_E \rangle$	0.3 nA, $\sigma = 20\%$	adaptation strength (RS), lognormal
$\langle \tau_{a,S} \rangle$	50 ms, $\sigma = 20\%$	adaptation time constant (SOM), lognormal
$\langle \Delta a_S \rangle$	0.2 nA, $\sigma = 20\%$	adaptation strength (SOM), lognormal

BCN network parameters

Symbol	Value	Description
N_E	2000	number of RS excitatory neurons in the BCN
N_I	400	number of FS inhibitory neurons in the BCN
N_S	200	number of SOM-LTS inhibitory neurons in the BCN

$R_{m,k}I_0$	10 mV	constant external input to BCN neurons
$r_{\text{ext,th}}$	10 Hz	rate of external "thalamic" inputs
$C_{\text{ext,th}}$	500	number of external "thalamic" inputs
$r_{\text{ext,bc}}$	2 Hz	rate of external "cortical" inputs
$C_{\text{ext,bc,e}}$	2000	number of external "cortical" inputs to RS neurons
$C_{\text{ext,bc,i}}$	1000	number of external "cortical" inputs to FS neurons
C_{ee}	300	synapses from RS to RS (15% connection prob.)
C_{ei}	200	synapses from FS to RS (50% connection prob.)
C_{es}	100	synapses from SOM to RS (50% connection prob.)
C_{ie}	800	synapses from RS to FS (40% connection prob.)
C_{ii}	200	synapses from FS to FS (50% connection prob.)
C_{is}	50	synapses from SOM to FS (25% connection prob.)
C_{se}	1000	synapses from RS to SOM (50% connection prob.)
C_{si}	100	synapses from FS to SOM (25% connection prob.)
C_{ss}	0	synapses from SOM to SOM (0% connection prob.)
$\langle J_{ee} \rangle$	0.1 mV	syn. strength from RS to RS, exponential
$\langle J_{ei} \rangle$	0.5 mV	syn. strength from FS to RS, exponential
$\langle J_{es} \rangle$	0.25 mV	syn. strength from SOM to RS, exponential
$\langle J_{ie} \rangle$	0.2 mV	syn. strength from RS to FS, exponential
$\langle J_{ii} \rangle$	1.0 mV	syn. strength from FS to FS, exponential
$\langle J_{is} \rangle$	0.1 mV	syn. strength from SOM to FS, exponential
$\langle J_{se} \rangle$	0.1 mV	syn. strength from RS to SOM, exponential
$\langle J_{si} \rangle$	0.25 mV	syn. strength from FS to SOM, exponential
$D_{\min}$ ($D_{\max}$)	0.5 ms (1.0 ms)	min. (max.) transmission delay, uniform
$\left\langle \hat{J}_{ii} \right\rangle$	0.05 mV	strength of gap junctions between FS neurons
$\left\langle \hat{J}_{ss} \right\rangle$	0.05 mV	strength of gap junctions between SOM neurons
$\hat{D}_{\min}$ ($\hat{D}_{\max}$)	0.1 ms (0.5 ms)	min. (max.) gap junction transmission delay, uniform

Short-term plasticity parameters

Symbol	Value	Description
$\tau_{D,s}$	150 ms	dep. time constant (strong STD)
$U_{se,s}$	0.2	synaptic release probability (strong STD)
$\tau_{D,w}$	50 ms	dep. time constant (weak STD)
$U_{se,w}$	0.05	synaptic release probability (weak STD)
$\tau_{D,f}$	100 ms	dep. time constant (strong STF)
τ_F	300 ms	fac. time constant (strong STF)
U	0.03	initial increase for facilitation variable u (strong STF)
U_b	0.01	resting value for facilitation variable u (strong STF)
τ_f	250 ms	syn. fail rate recovery time constant (RS-to-SOM)
Δp_f	0.1	fail rate decrease per spike (RS-to-SOM)
$p_{f,\text{rest}}$	0.5	syn. fail rate at rest (RS-to-SOM)
$p_{\min}$	0.1	min. syn. fail rate (RS-to-SOM)

Readout parameters

Symbol	Value	Description
$\hat{C}$	1000	size of readout set for IR and DR $\mathcal{S}^A$
ΔT	10 ms	time shift for DR
τ_f	15 ms	time constant for readout filter
N_B	10000	RS neurons in $\mathcal{S}^B$ (DNR)
$N_{\mathcal{I}}$	2500	FS neurons in $\mathcal{I}$ (DNR)
$R_{m,k}I_0$	15 mV	constant external input to DNR neurons
$C_{\text{ext,th}}$	250	number of external "thalamic" inputs to DNR neurons
$C_{\text{ext,bc,e}}$	2000	number of external "cortical" inputs to $\mathcal{S}^B$
$C_{\text{ext,bc,i}}$	1000	number of external "cortical" inputs to $\mathcal{I}$
$\hat{C}$	1000	feed-forward inputs per neuron to DNR
$\left\langle J_e^{FF} \right\rangle$	0.1 mV	syn. strength from BCN (RS only) to $\mathcal{S}^B$, exponential
$\left\langle J_i^{FF} \right\rangle$	0.2 mV	syn. strength from BCN (RS only) to $\mathcal{I}$, exponential

$\left\langle C_{ii}^{R} \right\rangle$	200	inputs from $\mathcal{I}$ to $\mathcal{I}$
$\left\langle C_{ei}^{R} \right\rangle$	200	inputs from $\mathcal{I}$ to $\mathcal{S}^{B}$
$\left\langle J_{ii}^{R} \right\rangle$	1.0 mV	syn. strength from $\mathcal{I}$ to $\mathcal{I}$, exponential
$\left\langle J_{ei}^{R} \right\rangle$	0.65 mV	syn. strength from $\mathcal{I}$ to $\mathcal{S}^{B}$, exponential
$D_{\min}^{FF}$ ($D_{\max}^{FF}$)	0.5 ms (1 ms)	min. (max.) delay from BCN to DNR, uniform
ΔT	10 ms	additional delay from BCN to $\mathcal{I}$

General parameters

Symbol	Value	Description
T_s	100 ms to 400 ms	stimulus duration
$R_m \Delta I_{\mathrm{max,e}}$	5 nA	maximum current intensity for RS $\mathcal{B}_0$
$R_m \Delta I_{\mathrm{max,i}}$	2.5 nA	maximum current intensity for FS $\mathcal{B}_0$
T_w	600 ms	time window for single-cell detection
T_{ic}	1200 ms	simulation time to forget initial conditions
T	1400 ms	simulation time (data acquisition)
Δt	0.1 ms	simulation time step
N_{trials}	10000	number of trials for each condition

Chapter 5.

Concluding Remarks

The rat barrel cortex is so sensitive to inputs that even the stimulation of a single neuron can provoke a weak but measurable effect (Houweling and Brecht, 2008). At the same time, cortical networks must be stable with respect to noise fluctuations to ensure a robust brain function. What properties of the network, encoding schemes, and readout principles make this balance of sensitivity and stability possible?

Merely scratching the surface of this problem of intimidating complexity, in this thesis the stability of the system was essentially taken for granted, by constructing network models that are known to produce a stable firing activity with characteristics compatible with those observed in the rat barrel cortex. The focus of the investigation was on devising readout schemes capable of detecting the effects of the single-cell stimulation in a robust and plausible way.

In chapter 2, the single-cell stimulation was replicated in a random network of leaky integrate-and-fire neurons tuned to fire in the asynchronous irregular state at low firing rates, and a first readout scheme to detect the stimulation was proposed. This detector considered the filtered activity of a subset of the network and reacted to deviations from the spontaneous activity. When the readout was sufficiently biased towards the neurons receiving direct input from the stimulated cell, the stimulation could be detected, and stimulating an inhibitory cell had a stronger effect for a given bias, as in the experiments by Houweling and Brecht (2008).

Analytical approximations to the detection rates were derived, which elucidated the link between model parameters, the network dynamics, and the detectability of the single-cell stimulation. This theoretical approximation emphasized the importance of cross-correlations as a limiting factor for the detectability. Because of their importance, cross-correlations were studied analytically in a linear-response framework. Although the final expression for the cross-spectrum turned out to be equivalent to a formula previously derived by Trousdale et al. (2012), the calculation (presented separately in appendix A) offers a complementary approach to the problem.

The findings of chapter 2 demonstrate that a preeminent role among network parameters is played by the strength of the network coupling, which profoundly affects both the transmission of the signal (the single-cell stimulation) and the intensity of the noise generated by the network. As a consequence, maximum detectability is achieved when the coupling strength is chosen in

an intermediate range.

In the experiments by Houweling and Brecht (2008), the training phase was a necessary condition for the single-cell stimulation to be detectable. The readout bias introduced in chapter 2 offers a first, simple proposal for how the system can be modified from its "natural" state to obtain results compatible with the experimental findings. In other words, it could represent the way animals learn their task. However, a shortcoming of this readout hypothesis is that the decision about the presence of the stimulation is taken locally, i.e. within the same network being stimulated, which is perhaps an unrealistic assumption in a primary sensory area such as the barrel cortex. To overcome this limitation, chapter 3 introduced a new detection scheme, in which a second network acted as a readout circuit. Notably, this new readout network proved to be not only more plausible, but also more effective in detecting the single-cell stimulation because it required a much smaller bias to detect the single-cell stimulation with the same reliability of the readout from chapter 2. Both simulations and a linear-response analysis made clear that the advantage of the new readout network of chapter 3 derives from its own population of inhibitory neurons. The feed-forward inhibition removes global cross-correlations without suppressing the signal, provided that connections from the stimulated network to the readout are slightly biased towards or against the direct neighbors of the stimulated cell.

Although the model of chapter 3 represented a clear improvement in terms of realism, there is still reason to doubt that the activation of a second network receiving direct input from the barrel cortex could suffice to determine the licking response because direct connections from the barrel cortex to the primary motor cortex exist, but they target the whisker-related areas, and not those controlling tongue movements (Alloway et al., 2004; Chakrabarti et al., 2008). Therefore, it would be an important question for future studies to extend the readout network to a third processing stage.

The results of further single-cell stimulation experiments by Brecht and coworkers investigating how the properties of the current injection influence the response probability provide an additional constraint that the model should satisfy. In short, a constant current produced an effect which was essentially independent of the duration and intensity of the stimulation, whereas an irregular stimulus was significantly more detectable (Doron, 2012; Doron et al., 2014). The effort made in chapter 4 to develop a model complying with these constraints proceeded along two lines. One approach was to include biological details of the barrel cortex into the recurrent network representing the stimulated brain region. The most important additions were short-term plasticity for all synaptic connections, spike-frequency adaptation, and heterogeneity in the cellular parameters and connection probabilities for the three types of cell included in the model (excitatory regular spiking cells, fast-spiking inhibitory interneurons, and low-threshold spiking somatostatin-positive inhibitory cells). The second strategy was to change the functioning principle of the readout from a leaky integrate-and-react prescription, used in chapters 2

and 3, to a differentiator readout, which reacted to variations of the input activity. This new readout prescription was tested by using an abstract mathematical implementation and then implemented by means of a network of integrate-and-fire neurons.

The core results of chapter 4 were that the differentiator readout can detect the stimulation of an excitatory regular-spiking neuron with a reliability that, as in the experiments, is essentially independent of the duration and intensity of a constant stimulus, but increases if an irregular stimulation is used. In the model, low-threshold spiking somatostatin-positive interneurons are essential for the detectability of excitatory cells. This hypothesis has already been formulated in the experimental literature (Doron and Brecht, 2015). The effect size that can be reached upon the stimulation of a fast-spiking inhibitory cell was somewhat smaller in the model than in the data, and it required a strong bias in the readout, which indicates that the usually stronger effect of fast-spiking cells observed experimentally is possibly based on a mechanism that is missing in the model.

As far as the dependencies on the parameters of the stimulation are concerned, the integrator readout yielded results that are in disagreement with the experiments. However, it detected the stimulation more reliably than the differentiator, which raises the question of why the animal should opt for the less efficient differentiator readout and lose rewards. A possible answer could be that the integrator is only more reliable in the stationary situation considered in this thesis, but it would be less efficient than the differentiator in the presence of non-stationary modulations of the spontaneous state, which are indeed present in the cortical activity of an animal that is awake (see fig. 4.21). It would be interesting to test this hypothesis by including a slow random modulation of the network activity to the model of chapter 4.

One limitation common to all models studied here is that the network topology was represented by a random and homogeneous graph, which is a reasonable approximation only on rather small scales. An important question is whether the main findings of this thesis hold in networks with a spatial structure, such as a distance-dependent connection probability or a partial clustering of connections representing, for instance, the layered structure of the cortex and the partial barrel segregation.

In all models of this thesis, it was always assumed that the training phase had already taken place. The effects of the training were modeled mainly by the readout bias and, in chapter 4, by the tuning of the readout network necessary to make it operate as a differentiator. It is an interesting question whether the same results can be obtained by applying a suitable learning rule. A considerable complication arises, however, from the fact that the training was performed by using microstimulation, which generates a complex cascade of events only partially understood.

Microstimulation pulses were also randomly delivered between single-cell stimulation trials to keep rats focused. Most experimental evidence associates desynchronized firing with the attentive state and slow oscillations with quiet wakefulness and sleep (Renart et al., 2010; Harris

and Thiele, 2011; Poulet et al., 2012). Hence, the network models of this thesis were tuned to fire in the asynchronous irregular regime. Nevertheless, oscillations are not completely absent in the awake organism, and it cannot be excluded that they play a role in the detectability of the single-cell stimulation, perhaps by inducing a sudden phase shift in an oscillation. An interesting question is how coherent the oscillation would have to be to permit this kind of detection and how the corresponding readout would need to work.

The scarcely constrained nature of the problem considered in this thesis grants a great deal of freedom in constructing a model. However, it is likely that advances in experimental techniques will soon permit to combine the single-cell stimulation with simultaneous and accurate recordings of the local network activity. This kind of experiments would certainly prove valuable as a source of inspiration and direction for future theoretical investigations.

Appendix A.

Linear Response Theory for Network Cross-Correlations

The goal of this appendix is to derive a theoretical approximation for the cross-spectrum between neurons of the recurrent network considered in chapter 2, i.e. a homogeneous random network of excitatory and inhibitory leaky integrate-and-fire neurons with fixed in-degree and exponentially distributed weights.

As mentioned in section 2.3.2, the theoretical characterization of cross-correlations in recurrent networks is by no means a closed research topic (Lindner et al., 2005; Ostojic et al., 2009a; Trousdale et al., 2012; Helias et al., 2013, 2014). The starting point of most studies is to consider the recurrent input as a small perturbation of the dynamics driven by the external noise. This core idea was put forward by Lindner et al. (2005), in which the linear-response ansatz in the Fourier domain was applied to a *single realization* of a spike-train and not just to its trial average:

$$\tilde{x}_k(f) = \tilde{x}_{0,k}(f) + \sum_j K_{kj}(f)\tilde{x}_j(f), \tag{A.1}$$

where $\tilde{x}_{0,k}(f)$ is the activity of neuron k in the absence of recurrent connections, i.e. driven only by eternal input, and $K_{kj}(f)$ describes the linear effect of neuron j on the firing rate of neuron k and contains the single-neuron susceptibility $\chi(f)$.

Trousdale et al. (2012) used the ansatz in eq. (A.1) to solve the problem "top-down": they started with a general expression involving the formal inversion of the interaction matrix and then, through a series expansion, found an explicit result for different types of network graphs, including a random network with fixed in-degree. However, their starting point, eq. (A.1), assumes a strong external drive. Because the external noise in the networks considered in chapter 2 is either not particularly strong or completely absent, an alternative "bottom-up" approach was attempted here, which does not explicitly assumes the presence of external noise. Rather, it assumes only that correlations are weak, so that contributions from different correlation sources can be described by linear equations and simply sum. Relating these contributions to each other and imposing self-consistency yields a solution depending only on the power spectrum of

the single neuron activity, which, as shown in section 2.1, can be roughly approximated by the theoretical expression for a neuron driven by excitatory and inhibitory shot noise (Richardson and Swarbrick, 2010).

Analytical approximations derived here are compared to the three main parameter sets used in chapter 2. The first is reported in table 2.1 and corresponds to the large network without external noise considered throughout chapter 2. For brevity, in the following this parameter set will be referred to as "standard autonomous network" (SAN). The second parameter set differs from the first only through the presence of external noise. This parameter set is defined in table 2.2 and will be called "standard driven network" (SDN). The last parameter set is the "single-barrel network" (SBN) considered in section 2.6 and listed in table 2.3.

A.1. Notation and general approach

Consider a pair of randomly selected neurons in the recurrent network. Correlations between their spike trains arise from two main sources: the effect of spikes generated by one of the two neurons on the other (both via direct connections and via longer paths), and the effect of input correlations, i.e. correlations in the spikes received by the two neurons but that are not fired by either of them.

To separate these two sources, a second population of neurons can be introduced, as depicted in fig. A.1: here, the original recurrent network is labeled as Ω_1, while neurons in the new population Ω_0 only receive input from Ω_1 but have no output connections. More precisely, each neuron in Ω_0 receives input from C_E excitatory and C_I inhibitory randomly selected neurons in Ω_1 with weight and transmission delay drawn from the same distribution as for the recurrent network. In other words, neurons in Ω_0 receive the same input as neurons in Ω_1 but their spikes do not affect any other part of the network. Because neurons within Ω_0 have no influence on each other, correlations between their spike trains can only be due to cross-correlations in their input. Let $S^{FF}_{x_1x_2}$ be the average cross-spectrum between them a pair of neurons in Ω_0, and $\eta_1(t)$ and $\eta_2(t)$ the total synaptic input to the first and second neuron, respectively. The Fourier transformed inputs to the two neurons, $\tilde{\eta}_1(f)$ and $\tilde{\eta}_2(f)$, will be approximated by two Gaussian noise processes. Each of these processes can be decomposed by means of three uncorrelated processes $\xi_c(f)$, $\xi_1(f)$, and $\xi_2(f)$

$$\tilde{\eta}_1(f) = \xi_c(f) + \xi_1(f) \tag{A.2}$$

and

$$\tilde{\eta}_2(f) = \xi_c(f) + \xi_2(f). \tag{A.3}$$

where the common process is defined so that its power spectrum is equal to the cross-spectrum

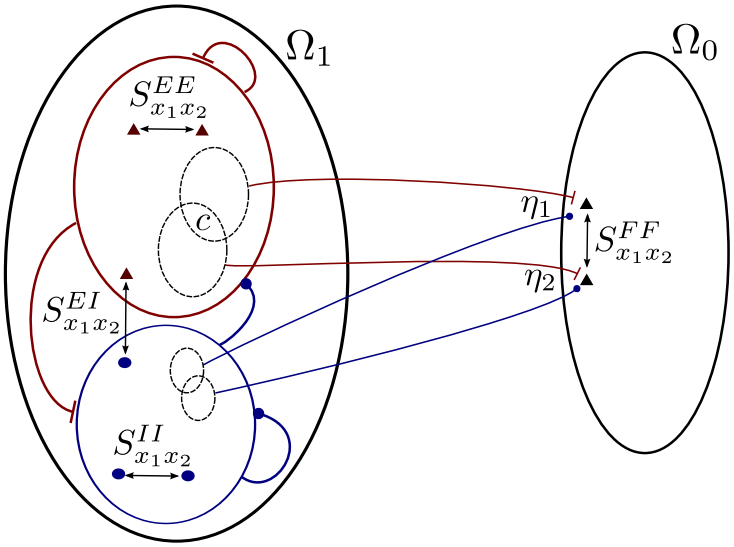

Figure A.1. – **Notation used in this appendix.** The recurrent network considered in this appendix is labeled as Ω_1. A second population Ω_0 consists of neurons receiving input connections from the recurrent network with the same topology, weights, and delays as neurons within the recurrent network but without any outgoing connections. In other words, neurons in Ω_0 receive input statistics (including input correlations) as if they were part of the recurrent network Ω_1, but cannot influence one another. The cross-spectrum between two neurons within Ω_0 is $S^{FF}_{x_1x_2}$. The input from Ω_1 to two generic neurons in Ω_0 is indicated by η_1 and η_2. On average, two neurons in Ω_0 share $c = p_c C_E$ excitatory and γc inhibitory inputs. The cross-spectrum between two excitatory neurons within Ω_1 is $S^{EE}_{x_1x_2}$, while $S^{EI}_{x_1x_2}$ and $S^{II}_{x_1x_2}$ are the cross-spectra for the other two possible cell type combinations.

between $\eta_1(t)$ and $\eta_2(t)$, i.e. $\langle \xi_c(f)\xi_c^*(f)\rangle = T \cdot S_{\eta_1\eta_2}(f)$ is satisfied (the total duration of the simulation, T, is assumed to be large enough so that the spectra do not appreciably depend on T). The two processes $\xi_{1,2}(f)$ are uncorrelated with each other and with ξ_c. Note that $\xi_c(f)$ represents not only the input spike trains common to the two neurons, but also global correlations in the recurrent network. In the following, the frequency dependence of all stochastic processes will be omitted for brevity. If ξ_c is considered to be a frozen signal and the average is taken over the uncorrelated noise, the output of each neuron is a linear response to ξ_c (considering only $f > 0$)

$$\langle \tilde{x}_{1,2}\rangle_{\xi_{1,2}} = \chi(f)\xi_c, \tag{A.4}$$

where $\chi(f)$ is the susceptibility of the single neuron, i.e. the Fourier transform of the linear response kernel (see section 1.3 p. 29). Note that eq. (A.4) would hold for any signal strength if the input were exactly Gaussian (Novikov's theorem). As $\eta_{1,2}(t)$ is actually a superposition of spike trains, the linear response can be expected to hold if ξ_c is sufficiently weak compared to $\xi_{1,2}$. Under the same assumption, eq. (A.4) can be exploited to derive the output cross spectrum

as a function of the input cross-spectrum (Vilela and Lindner, 2009; Ostojic et al., 2009a):

$$S^{FF}_{x_1x_2}(f) = \frac{1}{T}\langle \tilde{x}_1 \tilde{x}_2^* \rangle_{\xi_c,\xi_1,\xi_2} = \frac{1}{T}\langle \langle \tilde{x}_1 \rangle_{\xi_1} \langle \tilde{x}_2^* \rangle_{\xi_2} \rangle_{\xi_c} = \frac{1}{T}|\chi(f)|^2 \langle \xi_c \xi_c^* \rangle_{\xi_c} = |\chi(f)|^2 S_{\eta_1\eta_2}(f). \quad \text{(A.5)}$$

Equation (A.5) is the starting point for the calculation of $S_{x_1x_2}(f)$, the average cross spectrum between neurons within Ω_1, the final goal of this appendix. The calculation is rather lengthy and is distributed across several sections. The goal of section A.3 is to derive an expression relating $S_{\eta_1\eta_2}(f)$ to the cross-spectrum between neurons in Ω_1. Because neurons in the recurrent network can influence each other, the cross-spectrum between two neurons in Ω_1 depends on their "identity", i.e. whether they are excitatory or inhibitory. In other words, the three cross spectra $S^{EE}_{x_1x_2}(f)$, $S^{II}_{x_1x_2}(f)$, and $S^{EI}_{x_1x_2}(f)$, the average cross-spectrum between two excitatory neurons, two inhibitory neurons, and a mixed excitatory-inhibitory pair, respectively, must be distinguished (see fig. A.1). Section A.4 aims at finding an expression linking $S^{EE}_{x_1x_2}(f)$, $S^{II}_{x_1x_2}(f)$, and $S^{EI}_{x_1x_2}(f)$ to $S^{FF}_{x_1x_2}(f)$. In the last section, all these expressions are combined to obtain first $S^{FF}_{x_1x_2}(f)$ and then the final result $S_{x_1x_2}(f)$. Before this program can be carried out, however, an expression for the other factor in eq. (A.5), i.e. the susceptibility $\chi(f)$, is needed.

A.2. Susceptibility and test of linear-response ansatz

The result eq. (A.5) is of little use if the firing-rate susceptibility of the single neuron in response to an additive signal is not known. The analytical expressions derived by Richardson and Swarbrick (2010) for a leaky integrate-and-fire neuron driven by exponential shot-noise include the average firing rate, the power spectrum and the linear response function with respect to the *input rates* (and not to an additive input). Droste and Lindner (2017a) calculated the susceptibility to an input current with excitatory background shot noise but without inhibitory noise.

Here, an approximation for $\chi(f)$ will be obtained from the well-known relationship linking the so-called DC susceptibility $\chi(f \to 0)$ to the derivative of the stationary firing rate eq. (2.9) with respect to its mean input $\mu = R_m I_0$ (see section 1.3 p. 30):

$$\chi(f \to 0) = \frac{\mathrm{d}\phi_{sn}}{\mathrm{d}\mu} = r_{\mathrm{sp}}^2 \tau_m \int_0^{1/J} \mathrm{d}s\, Z_0^{-1}(s) \left[\frac{e^{s(\hat{v}_T - \mu)}}{1 - Js} - e^{s(\hat{v}_R - \mu)} \right], \tag{A.6}$$

where $Z_0^{-1}(s) = (1 - Js)^{\tau_m C_E r_{\mathrm{sp}}} (1 + Jgs)^{\tau_m \gamma C_E r_{\mathrm{sp}}}$. The frequency-dependent part of the susceptibility can be approximated in the same way as in section 2.2 for the time-dependent response to a step stimulus, i.e. by an exponential decay. Translating this approximation into the frequency domain yields

$$\chi(f) \approx \frac{\chi(0)}{1 - 2\pi i f \tau_m}. \tag{A.7}$$

To compute the susceptibility from eqs. (A.6) and (A.7), the spontaneous firing rate r_{sp} is required. As discussed in chapter 2, the self-consistent solution of eq. (2.9) provides a fairly good estimate of r_{sp}. However, the value of the DC susceptibility depends rather strongly on the value of the spontaneous firing rate r_{sp}, as it appears not only as a quadratic prefactor in eq. (A.6), but also in Z_0^{-1} where it is multiplied by the fairly large factors $\tau_m C_E$ and $\tau_m \gamma C_E$. Consequently, a small imprecision in the theoretical r_{sp} can have a sizable impact on the susceptibility. In the following, $\hat{\chi}$ will denote the semi-theoretical susceptibility computed according to eqs. (A.6) and (A.7) while using the value of r_{sp} measured from simulations. It will be compared to χ, the susceptibility computed by inserting the theoretical r_{sp} into eqs. (A.6) and (A.7).

Before proceeding further with the analysis of the network cross-correlations, it is wise to check how well eqs. (A.6) and (A.7) describe the actual susceptibility of neurons subject to the network noise, as well as the applicability of the linear-response ansatz eq. (A.5). To this end, two tests will be used. First, a numerical estimate of the susceptibility can be obtained by measuring the ratio

$$\chi_{\mathrm{num}}(f) = \frac{\langle x \eta^* \rangle}{\langle \eta \eta^* \rangle}. \tag{A.8}$$

Secondly, if the linear response ansatz eq. (A.5) is valid, the square of the numerical susceptibility $|\chi_{\mathrm{num}}(f)|^2$ should be approximately equal to the ratio $\langle x_1 x_2^* \rangle / \langle \eta_1 \eta_2^* \rangle$.

The results of these two tests for the three main parameter sets used in chapter 2 are shown in fig. A.2. The nine panels of fig. A.2 are organized as follows: the three parameter sets are shown column-wise (first column, SAN; second column, SDN; third column, SBN; definitions are found on p. 196); the first two rows pertain to the first test (real and imaginary part, respectively); the third row shows results of the second test.

To describe fig. A.2 in more detail, one can start by considering fig. A.2A, which shows the real part of the susceptibility $\chi_{\mathrm{num}}(f)$ for the SAN parameter set (black continuous line) together with the theoretical approximation $\chi(f)$ computed via eqs. (A.6) and (A.7) (red continuous line). While the global shape of the two curves is similar, the oscillatory features at higher frequencies are completely ignored by the exponential-decay assumption eq. (A.7). A more evident discrepancy is seen in the lower frequency range, which saturates at a lower level than predicted by the theory. As pointed out above, the low-frequency limit depends strongly on the spontaneous firing rate. In fact, $\hat{\chi}$, the theoretical susceptibility corrected by using the measured value r_{sp}, (red dashed line) is closer to the measured χ_{num}, but still overestimates the actual rate response. For the SDN parameters, however, the agreement is generally better (fig. A.2B): the theoretical χ is again higher than χ_{num} for low frequencies, but decays faster; the agreement of $\hat{\chi}$ (red dashed line) at low frequencies is rather good. The same qualitative picture is found for the SBN parameter set, shown in fig. A.2C. Here, the susceptibility is lower because of the strong external input.

The imaginary parts of the same quantities are shown in the second row of fig. A.2 with the same color and line coding. The shape of the imaginary part of χ is similar for all parameter sets, with a minimum around 10 Hz. The theory again overestimates the magnitude of the linear response, especially for the SAN case (fig. A.2D). The agreement is better for the SDN (fig. A.2E), and is best for the case of the SBN (fig. A.2F), for which the external noise is strongest.

The third row of fig. A.2 displays results from the second test. In the SAN (fig. A.2G), the absolute square of the susceptibility $|\chi_{\mathrm{num}}|^2$, plotted in black, exceeds the actual response ratio $\langle x_1 x_2^* \rangle / \langle \eta_1 \eta_2^* \rangle$ (green line) by almost a factor two. Consequently, the discrepancy with the theoretical approximation $|\chi|^2$ ($|\hat{\chi}|^2$), plotted as the continuous (dashed) red line for the theoretical (measured) r_{sp}, is even larger. In the SDN, however, the ratio $\langle x_1 x_2^* \rangle / \langle \eta_1 \eta_2^* \rangle$ (green line) is rather close to $|\chi_{\mathrm{num}}|^2$ (black line) and to $|\hat{\chi}|^2$, (red dashed line), but somewhat below $|\chi|^2$. For the SBN (fig. A.2I) the picture is similar. Here, however, the ratio $\langle x_1 x_2^* \rangle / \langle \eta_1 \eta_2^* \rangle$ slightly exceeds $|\chi_{\mathrm{num}}|^2$ and is closer to $|\chi|^2$.

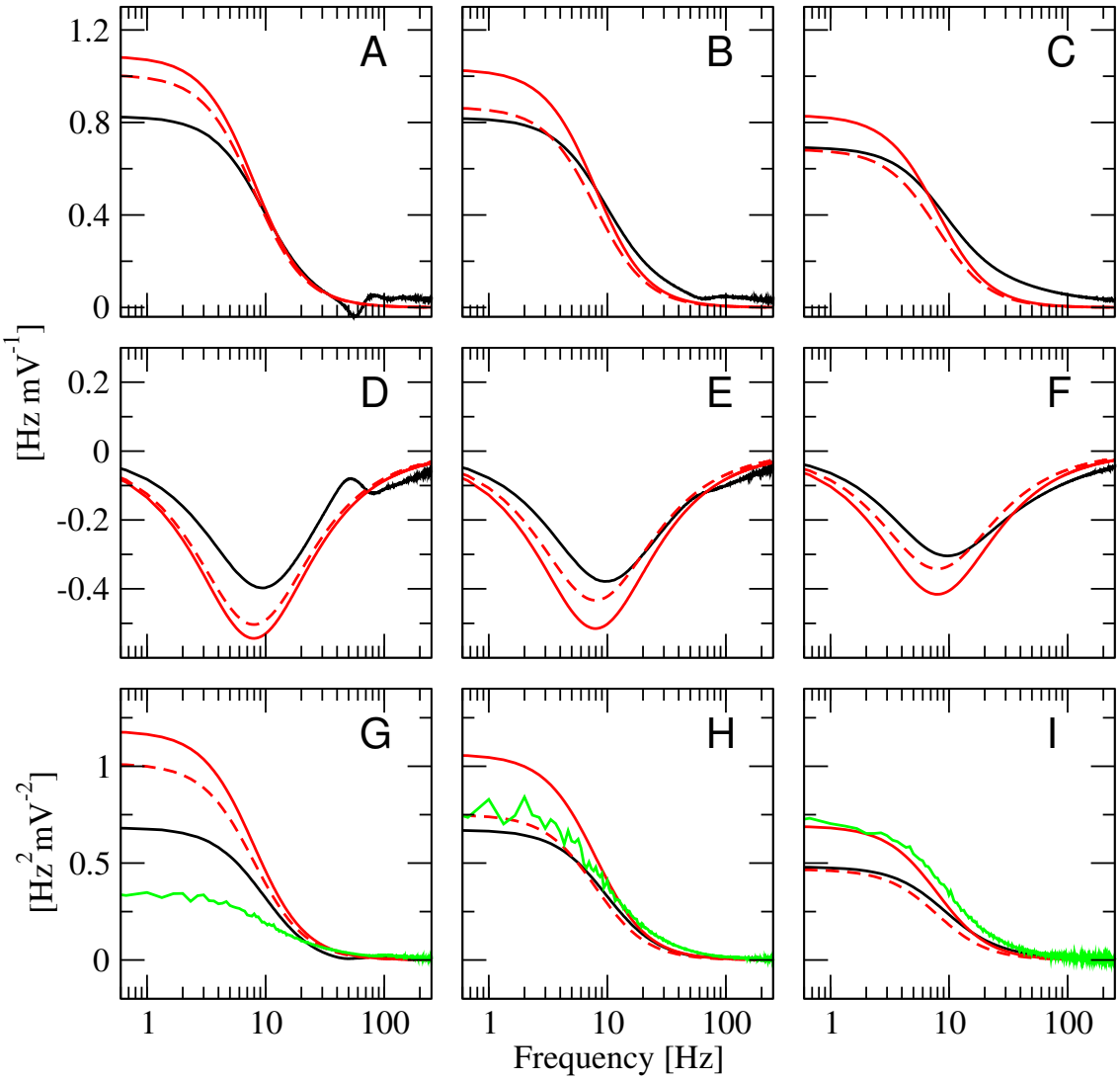

Figure A.2. – **Numerical tests for the approximation to the susceptibility.** First row (**A** to **C**): black line, real part of the numerically measured $\chi_{\rm num}(f)$ defined in eq. (A.8); red line, real part of $\chi(f)$ defined in eq. (A.7); red dashed line, real part of $\hat{\chi}$, the theoretical susceptibility corrected with the measured $r_{\rm sp}$. Second row, (**D** to **F**): black line, imaginary part of $\chi_{\rm num}(f)$; red continuous line, $\chi(f)$; red dashed line, $\hat{\chi}(f)$. Third row, (**G** to **I**): black line, $|\chi_{\rm num}(f)|^2$; red continuous, $|\chi(f)|^2$; red dashed line, $|\hat{\chi}(f)|^2$; green line, $\langle x_1 x_2^* \rangle / \langle \eta_1 \eta_2^* \rangle$. Parameters (**A**,**D**,**G**): "standard autonomous network" table 2.1; (**B**,**E**,**H**): "standard driven network" table 2.2; (**C**,**F**,**I**): "single-barrel network" table 2.3. Averages are performed over 4000 neuron pairs and 2000 network realizations.

A.3. Cross-spectrum of input currents

The goal of this section is to derive an expression relating the input cross-spectrum $S_{\eta_1\eta_2}(f)$ to the single neuron power spectrum $S_{xx}(f)$, and to the three types of cross-spectra between spike trains in the recurrent network, i.e. $S^{EE}_{x_1x_2}(f)$, $S^{EI}_{x_1x_2}(f)$, and $S^{II}_{x_1x_2}(f)$. To simplify the notation, all tildes indicating Fourier transformation as well as the frequency argument of spike trains and of power- and cross-spectra will be dropped from this point on.

Consider two neurons within Ω_0. The input to any of the two neurons (here $a = 1, 2$) is given by:

$$\eta_a = \tau_m \left(\sum_i^{C_E} J_{ap_{a,i}} x^E_{p_{a,i}} e^{2\pi i f D_{ap_{a,i}}} - g \sum_j^{C_I} J_{aq_{a,j}} x^I_{q_{a,j}} e^{2\pi i f D_{aq_{a,j}}} \right), \tag{A.9}$$

where $\{p_{a,i}\}_{i=1,2\ldots C_E}$ runs over the indices of all excitatory presynaptic neurons of a, and $\{q_{a,j}\}_{j=1,2\ldots C_I}$ runs over the indices of all inhibitory presynaptic neurons of a. Because the presynaptic neurons are chosen at random, these two sets of indices have a random overlap. In other words, each neuron pair shares c_E excitatory and c_I inhibitory inputs, where c_E and c_I follow hypergeometric distributions[1] with means $c = p_c C_E$ and γc, respectively. If the contribution of the shared inputs to η_1 is separated from the rest, the total input to the first neuron can be written as

$$\eta_1 = \eta_{c,1} + \eta_{u,1}, \tag{A.10}$$

where $\eta_{c,1}$ is the input from the c_E neurons that are connected to both neurons, and $\eta_{u,1}$ is the input from neurons that are connected to the first neuron but not to the second. Analogously,

$$\eta_2 = \eta_{c,2} + \eta_{u,2}. \tag{A.11}$$

The explicit expressions for the shared input terms read:

$$\eta_{c,1} = \tau_m \left(\sum_i^{c_E} J_{1k_i} e^{2\pi i f D_{1k_i}} x^E_{k_i} - g \sum_j^{c_I} J_{1\ell_j} e^{2\pi i f D_{1\ell_j}} x^I_{\ell_j} \right) \tag{A.12}$$

and

$$\eta_{c,2} = \tau_m \left(\sum_i^{c_E} J_{2k_i} e^{2\pi i f D_{2k_i}} x^E_{k_i} - g \sum_j^{c_I} J_{2\ell_j} e^{2\pi i f D_{2\ell_j}} x^I_{\ell_j} \right). \tag{A.13}$$

[1] It is the distribution of c_E (c_I) hits in C_E (C_I) draws without replacement in a population of N_E (N_I) neurons. It is very similar to a binomial distribution when the population is much larger than the number of draws.

The non-shared input is:

$$\eta_{u,1} = \tau_m \left(\sum_i^{c'_E} J_{1m_{1,i}} e^{2\pi i f D_{1m_{1,i}}} x^E_{m_{1,i}} - g \sum_j^{c'_I} J_{1n_{1,j}} e^{2\pi i f D_{1n_{1,j}}} x^I_{n_{1,j}} \right) \tag{A.14}$$

and

$$\eta_{u,2} = \tau_m \left(\sum_i^{c'_E} J_{2m_{2,i}} e^{2\pi i f D_{2m_{2,i}}} x^E_{m_{2,i}} - g \sum_j^{c'_I} J_{2n_{2,j}} e^{2\pi i f D_{2n_{2,j}}} x^I_{n_{2,j}} \right) \tag{A.15}$$

where $c'_E = C_E - c_E$ and $c'_I = C_I - c_I$. Note that in $\eta_{c,1}$ and $\eta_{c,2}$ the same spike trains appear with different synaptic amplitudes and delays, whereas in $\eta_{u,1}$ and $\eta_{u,2}$ input spike trains are also different (and non-overlapping by definition). The cross-spectrum $S_{\eta_1\eta_2}$ can be split into four contributions:

$$TS_{\eta_1\eta_2} = \langle\langle \eta_1 \eta_2^* \rangle\rangle_{\text{net}} = \left\langle \left\langle \eta_{c,1}\eta^*_{c,2} \right\rangle \right\rangle_{\text{net}} + \left\langle \left\langle \eta_{c,1}\eta^*_{u,2} \right\rangle \right\rangle_{\text{net}} + \left\langle \left\langle \eta_{u,1}\eta^*_{c,2} \right\rangle \right\rangle_{\text{net}} + \left\langle \left\langle \eta_{u,1}\eta^*_{u,2} \right\rangle \right\rangle_{\text{net}}. \tag{A.16}$$

In eq. (A.16), the angular brackets without index $\langle\cdot\rangle$ indicate an average over trials (and noise realizations), and $\langle\cdot\rangle_{\text{net}}$ denote averaging over distinct neuron pairs or different network realizations. In the following, the stochastic overlaps between the two excitatory and the two inhibitory presynaptic populations of the considered neuron pair, indicated by c_E and c_I, respectively, appear as summation limits. Because c_E and c_I depend on the particular pair and on the network realization, it will be necessary to split the averaging over networks as follows:

$$\langle\cdot\rangle_{\text{net}} = \left\langle \langle\cdot\rangle_{\text{net}|\{c_E,c_I\}} \right\rangle_{\{c_E,c_I\}}, \tag{A.17}$$

where the inner angular brackets indicate an average over all distinct neuron pairs and network realizations, in which the value of c_E and c_I is fixed, and the outer brackets indicate an average over the distribution of $\{c_E, c_I\}$. With these definitions, each term can be calculated explicitly by inserting eqs. (A.12) to (A.15) into eq. (A.16). The first term reads

$$\begin{aligned} \left\langle \left\langle \eta_{c,1}\eta^*_{c,2} \right\rangle \right\rangle_{\text{net}} = \tau_m^2 \Bigg\langle & \sum_{i,j}^{c_E,c_E} J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2k_j} e^{-2\pi i f D_{2k_j}} \left\langle x^E_{k_i} x^{E*}_{k_j} \right\rangle \\ & - g \sum_{i,j}^{c_E,c_I} J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2\ell_j} e^{-2\pi i f D_{2\ell_j}} \left\langle x^E_{k_i} x^{I*}_{\ell_j} \right\rangle \\ & - g \sum_{i,j}^{c_I,c_E} J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2k_j} e^{-2\pi i f D_{2k_j}} \left\langle x^I_{\ell_i} x^{E*}_{k_j} \right\rangle \\ & + g^2 \sum_{i,j}^{c_I,c_I} J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2\ell_j} e^{-2\pi i f D_{2\ell_j}} \left\langle x^I_{\ell_i} x^{I*}_{\ell_j} \right\rangle \Bigg\rangle_{\text{net}} \end{aligned} \tag{A.18}$$

$$
\begin{aligned}
&= \tau_m^2 \Bigg\langle \Bigg\langle \sum_{i=j}^{c_E} J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2k_i} e^{-2\pi i f D_{2k_i}} \left\langle x_{k_i}^E x_{k_i}^{E*} \right\rangle \Bigg\rangle_{\text{net}|\{c_E,c_I\}} \qquad \text{(A.19)} \\
&\quad + g^2 \Bigg\langle \sum_{i=j}^{c_I} J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2\ell_i} e^{-2\pi i f D_{2\ell_i}} \left\langle x_{\ell_i}^I x_{\ell_i}^{I*} \right\rangle \Bigg\rangle_{\text{net}|\{c_E,c_I\}} \\
&+ \Bigg\langle \sum_{i\neq j}^{c_E,c_E} J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2k_j} e^{-2\pi i f D_{2k_j}} \left\langle x_{k_i}^E x_{k_j}^{E*} \right\rangle \Bigg\rangle_{\text{net}|\{c_E,c_I\}} \\
&- g \Bigg\langle \sum_{i,j}^{c_E,c_I} \Big[J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2\ell_j} e^{-2\pi i f D_{2\ell_j}} \left\langle x_{k_i}^E x_{\ell_j}^{I*} \right\rangle \\
&\qquad\qquad + J_{1\ell_j} e^{2\pi i f D_{1\ell_j}} J_{2k_i} e^{-2\pi i f D_{2k_i}} \left\langle x_{k_i}^E x_{\ell_j}^{I*} \right\rangle^* \Big] \Bigg\rangle_{\text{net}|\{c_E,c_I\}} \\
&+ g^2 \Bigg\langle \sum_{i\neq j}^{c_I,c_I} J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2\ell_j} e^{-2\pi i f D_{2\ell_j}} \left\langle x_{\ell_i}^I x_{\ell_j}^{I*} \right\rangle \Bigg\rangle_{\text{net}|\{c_E,c_I\}} \Bigg\rangle_{\{c_E,c_I\}} \\
&= \tau_m^2 \Bigg\langle \sum_{i=j}^{c_E} \left\langle J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2k_i} e^{-2\pi i f D_{2k_i}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{k_i}^E x_{k_i}^{E*} \right\rangle \right\rangle_{\text{net}|\{c_E,c_I\}} \qquad \text{(A.20)} \\
&\quad + g^2 \sum_{i=j}^{c_I} \left\langle J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2\ell_i} e^{-2\pi i f D_{2\ell_i}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{\ell_i}^I x_{\ell_i}^{I*} \right\rangle \right\rangle_{\text{net}|\{c_E,c_I\}} \\
&+ \sum_{i\neq j}^{c_E,c_E} \left\langle J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2k_j} e^{-2\pi i f D_{2k_j}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{k_i}^E x_{k_j}^{E*} \right\rangle \right\rangle_{\text{net}|\{c_E,c_I\}} \\
&- g \sum_{i,j}^{c_E,c_I} \Big[\left\langle J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2\ell_j} e^{-2\pi i f D_{2\ell_j}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{k_i}^E x_{\ell_j}^{I*} \right\rangle \right\rangle_{\text{net}|\{c_E,c_I\}} \\
&\qquad\qquad + \left\langle J_{1\ell_j} e^{2\pi i f D_{1\ell_j}} J_{2k_i} e^{-2\pi i f D_{2k_i}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{k_i}^E x_{\ell_j}^{I*} \right\rangle \right\rangle^*_{\text{net}|\{c_E,c_I\}} \Big] \\
&+ g^2 \sum_{i\neq j}^{c_I,c_I} \left\langle J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2\ell_j} e^{-2\pi i f D_{2\ell_j}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{\ell_i}^I x_{\ell_j}^{I*} \right\rangle \right\rangle_{\text{net}|\{c_E,c_I\}} \Bigg\rangle_{\{c_E,c_I\}} \\
&= T\tau_m^2 J^2 |\mathcal{D}(f)|^2 \Big\langle (c_E + g^2 c_I) S_{xx} + (c_E^2 - c_E) S_{x_1x_2}^{EE} \qquad \text{(A.21)} \\
&\qquad\qquad - 2g c_E c_I \Re[S_{x_1x_2}^{EI}] + g^2 (c_I^2 - c_I) S_{x_1x_2}^{II} \Big\rangle_{\{c_E,c_I\}}
\end{aligned}
$$

In eq. (A.19), diagonal and non-diagonal terms were separated. In the following step, eq. (A.20), the average over networks was applied to each term in each sum, which is possible because the value of each pair $\{c_E, c_I\}$ is frozen in the average. Furthermore, independence between J_{1j} and x_j and between J_{2i} and x_i is assumed, i.e. it is assumed that the activity of the j-th neuron is uncorrelated with the strength of one of its *outgoing* synapses. As autapses (self-connections) are not possible in the considered setup (i.e. $J_{ii} = 0$), the only possibility for J_{1j} to influence the activity of neuron j is through multiple synapses, i.e. over loops of length two or longer. Even if

the rare occurrence of a strong recurrent loop is unlikely to have a tangible effect on the summed input from many neurons considered here, this possibility cannot be completely ruled out *a priori* and the validity of this assumption is further discussed at the end of this section. In the diagonal sum terms, J_{1i} and J_{2i} are independent (as well as D_{1i} and D_{2i}) because connections from the same neuron to different targets are drawn independently, as all other weights and delays are. In the step from eq. (A.20) to eq. (A.21), each of the six sums is manipulated as follows

$$\tau_m^2 \left\langle \sum_{i=j}^{c_E} \left\langle J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2k_i} e^{-2\pi i f D_{2k_i}} \right\rangle_{\text{net}|\{c_E,c_I\}} \left\langle \left\langle x_{k_i}^E x_{k_i}^{E*} \right\rangle \right\rangle_{\text{net}|\{c_E,c_I\}} + \dots \right\rangle_{\{c_E,c_I\}} \tag{A.22}$$

$$= \tau_m^2 \left\langle \sum_{i=j}^{c_E} J^2 |\mathcal{D}(f)|^2 S_{xx} + \dots \right\rangle_{\{c_E,c_I\}} \tag{A.23}$$

$$= \tau_m^2 \left\langle J^2 |\mathcal{D}(f)|^2 S_{xx} \sum_{i=j}^{c_E} + \dots \right\rangle_{\{c_E,c_I\}} \tag{A.24}$$

$$= \tau_m^2 \left\langle J^2 |\mathcal{D}(f)|^2 S_{xx} c_E + \dots \right\rangle_{\{c_E,c_I\}} . \tag{A.25}$$

In eq. (A.22), the factors in the summand are all independent of each other by construction of the network. Furthermore, the average synaptic weight does not depend on $\{c_E, c_I\}$, i.e. $\langle J_{ij} \rangle_{\text{net}|\{c_E,c_I\}} = \langle J_{ij} \rangle_{\text{net}} = J$. The same holds for the delay terms $\langle e^{2\pi i f D_{ij}} \rangle_{\text{net}|\{c_E,c_I\}} = \langle e^{2\pi i f D_{ij}} \rangle_{\text{net}}$. The explicit expression of the characteristic function of the delay distribution, $\langle e^{2\pi i f D_{ij}} \rangle_{\text{net}} = \mathcal{D}(f)$ is given below. In eq. (A.25), the average over the distribution of $\{c_E, c_I\}$ could be performed by using known formulas for the hypergeometric distributions, as c_E and c_I are independent. However, several terms in eq. (A.21) will simplify in the final result, so that it is convenient to leave the averaging brackets as indicated.

The second term in eq. (A.16) can be calculated in the same way as above:

$$\begin{aligned}
\left\langle \left\langle \eta_{c,1} \eta_{u,2}^* \right\rangle \right\rangle_{\text{net}} = \tau_m^2 \Bigg\langle & \sum_{i,j}^{c_E,c_E'} J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2m_{2,j}} e^{-2\pi i f D_{2m_{2,j}}} \left\langle x_{k_i}^E x_{m_{2,j}}^{E*} \right\rangle \\
& - g \sum_{i,j}^{c_E,c_I'} J_{1k_i} e^{2\pi i f D_{1k_i}} J_{2n_{2,j}} e^{-2\pi i f D_{2n_{2,j}}} \left\langle x_{k_i}^E x_{n_{2,j}}^{I*} \right\rangle \\
& - g \sum_{i,j}^{c_I,c_E'} J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2m_{2,j}} e^{-2\pi i f D_{2m_{2,j}}} \left\langle x_{\ell_i}^I x_{m_{2,j}}^{E*} \right\rangle \\
& + g^2 \sum_{i,j}^{c_I,c_I'} J_{1\ell_i} e^{2\pi i f D_{1\ell_i}} J_{2n_{2,j}} e^{-2\pi i f D_{2n_{2,j}}} \left\langle x_{\ell_i}^I x_{n_{2,j}}^{I*} \right\rangle \Bigg\rangle_{\text{net}} \\
= T \tau_m^2 J^2 |\mathcal{D}(f)|^2 & \left\langle c_E c_E' S_{x_1 x_2}^{EE} - g c_E c_I' S_{x_1 x_2}^{EI} - g c_I c_E' S_{x_1 x_2}^{EI*} + g^2 c_I c_I' S_{x_1 x_2}^{II} \right\rangle_{\{c_E,c_I\}}
\end{aligned} \tag{A.26}$$

The third term in eq. (A.16) is - by symmetry - the complex conjugate of the second one.

$$\left\langle\left\langle \eta_{u,1}\eta^*_{c,2}\right\rangle\right\rangle_{\text{net}} = \left\langle\left\langle \eta_{u,2}\eta^*_{c,1}\right\rangle\right\rangle_{\text{net}} = \left\langle\left\langle (\eta_{c,1}\eta^*_{u,2})^*\right\rangle\right\rangle_{\text{net}} \tag{A.27}$$

The fourth and final term reads

$$\begin{aligned}
\left\langle\left\langle \eta_{u,1}\eta^*_{u,2}\right\rangle\right\rangle_{\text{net}} &= \tau_m^2 \Bigg\langle \sum_{i,j}^{c'_E,c'_E} J_{1m_{1,i}} e^{2\pi i f D_{1m_{1,i}}} J_{2m_{2,j}} e^{-2\pi i f D_{2m_{2,j}}} \left\langle x^E_{m_{1,i}} x^{E*}_{m_{2,j}}\right\rangle \\
&\quad - g \sum_{i,j}^{c'_E,c'_I} J_{1m_{1,i}} e^{2\pi i f D_{1m_{1,i}}} J_{2n_{2,j}} e^{-2\pi i f D_{2n_{2,j}}} \left\langle x^E_{m_{1,i}} x^{I*}_{n_{2,j}}\right\rangle \\
&\quad - g \sum_{i,j}^{c'_I,c'_E} J_{1n_{1,i}} e^{-2\pi i f D_{1n_{1,i}}} J_{2m_{2,j}} e^{-2\pi i f D_{2m_{2,j}}} \left\langle x^I_{n_{1,i}} x^{E*}_{m_{2,j}}\right\rangle \\
&\quad + g^2 \sum_{i,j}^{c'_I,c'_I} J_{1n_{1,i}} e^{-2\pi i f D_{1n_{1,i}}} J_{2n_{2,j}} e^{-2\pi i f D_{2n_{2,j}}} \left\langle x^I_{n_{1,i}} x^{I*}_{n_{2,j}}\right\rangle \Bigg\rangle_{\text{net}} \\
&= T\tau_m^2 J^2 |\mathcal{D}(f)|^2 \left\langle c'_E c'_E S^{EE}_{x_1x_2} - g c'_E c'_I S^{EI}_{x_1x_2} - g c'_I c'_E S^{EI*}_{x_1x_2} + g^2 c'_I c'_I S^{II}_{x_1x_2}\right\rangle_{\{c_E,c_I\}}
\end{aligned} \tag{A.28}$$

Adding the four terms together and dividing by the total time T yields the final result:

$$\begin{aligned}
S_{\eta_1\eta_2} &= \tau_m^2 J^2 |\mathcal{D}(f)|^2 \Bigg(\left\langle c_E + g^2 c_I\right\rangle_{\{c_E,c_I\}} S_{xx} \\
&\qquad + S^{EE}_{x_1x_2} \left\langle c_E^2 - c_E + 2c_E c'_E + c'_E c'_E\right\rangle_{\{c_E,c_I\}} \\
&\qquad - 2g\Re[S^{EI}_{x_1x_2}] \left\langle c_E c'_I + c_I c'_E + c_I c_E + c'_E c'_I\right\rangle_{\{c_E,c_I\}} \\
&\qquad + g^2 S^{II}_{x_1x_2} \left\langle c_I^2 - c_I + 2c_I c'_I + c'_I c'_I\right\rangle_{\{c_E,c_I\}} \Bigg) \\
&= \tau_m^2 J^2 \frac{2 - 2\cos(2\pi f\Delta)}{(2\pi f\Delta)^2} \Bigg(c(1+g^2\gamma) S_{xx} \\
&\qquad + S^{EE}_{x_1x_2}(C_E^2 - c) - 2g\gamma C_E^2 \Re[S^{EI}_{x_1x_2}] + g^2 S^{II}_{x_1x_2}(\gamma^2 C_E^2 - \gamma c)\Bigg)
\end{aligned} \tag{A.29}$$

where the explicit expression of the characteristic function of the delay distribution was inserted

$$|\mathcal{D}(f)|^2 = \left|\frac{1}{\Delta}\int_{D_{\min}}^{D_{\max}} \mathrm{d}D e^{2\pi i f D}\right|^2 = \left|\frac{e^{2\pi i f D_{\max}}(1 - e^{-2\pi i f\Delta})}{2\pi i f\Delta}\right|^2 = \frac{2 - 2\cos(2\pi f\Delta)}{(2\pi f\Delta)^2}, \tag{A.30}$$

and $\Delta = D_{\max} - D_{\min}$ is the width of the delay distribution. Intuitively, the effect of delays is

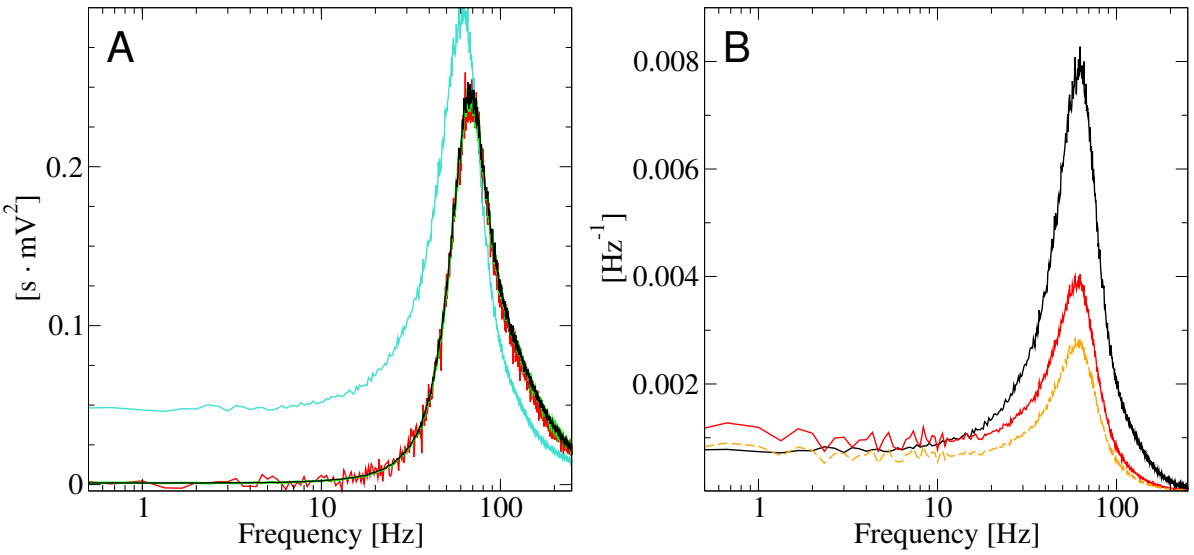

Figure A.3. – Comparison of eq. (A.29) to the cross-spectrum between input currents measured from simulation, and test of linear-response ansatz eq. (A.5). A: Input cross-spectrum $S_{\eta_1\eta_2}$ measured from simulation (black line); the right side of eq. (A.29) (red line); cross-spectrum between inputs generated by adding C_E and C_I randomly selected spike trains from the network with random weights and delays (green line); right side of eq. (A.29) with the identification $S^{EE}_{x_1x_2} = S^{EI}_{x_1x_2} = S^{II}_{x_1x_2} = S_{x_1x_2}$, where $S_{x_1x_2}$ is the average cross-spectrum between spike trains in the recurrent network (cyan). **B**: Comparison between average spike-train cross-spectrum $S^{FF}_{x_1x_2}$ measured from network simulation (black) and the combination of eq. (A.5) with eq. (A.29) (red continuous for theoretical χ and and dashed line for firing-rate corrected $\hat{\chi}$). Parameters: SDN (as in table 2.2).

negligible at low frequencies and $\mathcal{D}(f \to 0) \to 1$. For higher frequencies, if $\Delta \neq 0$ the delays disrupt cross-correlations as $\mathcal{D}(f \to \infty) \to 0$. The width of the distribution determines the rate of decay for $\mathcal{D}(f)$, while the mean delay does not matter. In fact, $\mathcal{D}(f)$ is due to the *difference* in the transmission delay from any source neuron to each of the two neurons. Adding the same delay to all connections, i.e. shifting the entire delay distribution, would not change these differences. Therefore, only the width of the distribution appears in eq. (A.30).

The validity of eq. (A.29) is checked in fig. A.3A for the SDN (other cases are qualitatively similar). First, the cross-spectrum $S_{\eta_1\eta_2}$ measured in simulations (black line) agrees well with the right side of eq. (A.29) (red line), where the three spike-train cross-spectra, $S^{EE}_{x_1x_2}$, $S^{EI}_{x_1x_2}$, and $S^{II}_{x_1x_2}$, are also measured from network simulations. Secondly, the assumption that spike trains are uncorrelated with the weight and delay of their outgoing synapses can be explicitly tested. To this end, two "surrogate" inputs $\hat{\eta}_1$ and $\hat{\eta}_2$ are generated with the following procedure: C_E excitatory and C_I inhibitory spike trains are randomly selected from the network, each spike train is multiplied with a random exponentially distributed weight and shifted by a random

delay. To mimic the common input, c excitatory and γc inhibitory spike-trains enter both $\hat{\eta}_1$ and $\hat{\eta}_2$ with a different weight and delay. In other words, $\hat{\eta}_1$ and $\hat{\eta}_2$ have the same statistics as η_1 and η_2 except for the correlation between each spike train and the prefactor through which it enters the sums in eqs. (A.21) and (A.26) to (A.28), which are prevented by construction. The cross-spectrum between these "artificially decorrelated" network inputs $S_{\hat{\eta}_1\hat{\eta}_2} = \langle \hat{\eta}_1 \hat{\eta}_2^* \rangle / T$ is shown as a green line in fig. A.3A and matches almost perfectly the actual cross-spectrum, which confirms the validity of the assumption made in the derivation of eq. (A.29). Finally, the consequences of ignoring the heterogeneity of the three spike-train cross-spectra are shown: the cyan line in fig. A.3A is obtained by setting $S^{EE}_{x_1x_2} = S^{EI}_{x_1x_2} = S^{II}_{x_1x_2} = S_{x_1x_2}$ in eq. (A.29), where $S_{x_1x_2}$ is the average cross-spectrum regardless of the neuron type. The result deviates by almost one order of magnitude in the low-frequency range. This discrepancy has the same source as the large numerical fluctuations seen for the correct theory plotted in red: the large prefactors (of which one is negative) multiplying the three different spectra.

If eq. (A.29) is inserted into eq. (A.5), an approximate expression for $S^{FF}_{x_1x_2}$ as a function of $S^{EE}_{x_1x_2}$, $S^{EI}_{x_1x_2}$, and $S^{II}_{x_1x_2}$ can be obtained. Figure A.3B demonstrates that the agreement between $S^{FF}_{x_1x_2}$ (black line) and $|\chi|^2 S_{\eta_1\eta_2}$ (red line) is reasonable for frequencies up to about 20 Hz. Using the firing-rate corrected $|\hat{\chi}|^2$ (dashed line) improves the agreement for low frequencies. Higher frequencies are strongly underestimated, as consequence of the crude approximation for the susceptibility in eq. (A.7).

A.4. Source of heterogeneity of spike-train cross-spectra in the recurrent network

The results of the previous section make clear that the differences between $S^{EE}_{x_1x_2}$, $S^{EI}_{x_1x_2}$, and $S^{II}_{x_1x_2}$ are essential to obtain the correct input cross spectrum. From where do these difference arise? As far as their input is concerned, all neuron pairs in Ω_1 are equivalent to each other and to the neurons in Ω_0. The only difference between a pair of neurons in Ω_1 and in Ω_0 are their output connections. Hence, the origin of the heterogeneity of the three spectra must lie in the spikes fired by the two neurons themselves. Because the effect of these spikes on the rest of the network depends on the type of the two neurons, the cross-spectrum between them is different for each possible combination. It is sufficient to consider three of the four cases, because $S^{EI}_{x_1x_2} = (S^{IE}_{x_1x_2})^*$. The simplest possibility for the neurons to influence each other is to be connected, either mono- or bidirectionally. If the two neurons are not directly connected, spikes from the first neuron can still affect the other (or itself) via an intermediate neuron. In a recurrent network, there are paths of arbitrary length connecting the two neurons. The effect of paths of increasing length is expected to decrease. However, the number of possible paths grows, which makes their combined effect possibly non-negligible. In the following, the effect of

paths of different lengths will be considered separately with the assumption that these effects do not interact with each other.

Impact of direct connections

The effect of a direct connection on the cross-spectrum between a pair of excitatory neurons will be examined first. Suppose that neuron 1 is connected to neuron 2 (but not vice-versa); then

$$x_2 \approx x_{2,0} + \chi\tau_m J_{21} e^{i2\pi f D_{21}} x_1 \tag{A.31}$$

where $x_{2,0}$ is the activity of the second neuron if the direct connection $1 \to 2$ is removed and higher order terms have been neglected. The linear-response assumption seems justified because of the large number of total inputs to neuron 2. Let $S^{EE}_{x_1x_2,1\to2}$ be the average cross-spectrum between excitatory neuron pairs in which a unidirectional connection from the first neuron to the second exists, but no connection in the opposite direction exists. Substituting eq. (A.31) into the definition of the cross-spectrum yields

$$S^{EE}_{x_1x_2,1\to2} \approx \frac{1}{T}\langle x_1 x^*_{2,0}\rangle + \frac{1}{T}\chi^*\tau_m J\mathcal{D}^*(f)\langle x_1 x_1^*\rangle = S^{EE}_{x_1x_2,nc} + \chi^*\tau_m J\mathcal{D}^*(f)S_{xx}, \tag{A.32}$$

where $\langle x_1 x^*_{2,0}\rangle/T = S^{EE}_{x_1x_2,nc}$ is the cross-spectrum between two excitatory neurons that are *not* directly connected to each other. It is convenient to introduce

$$A(f) = \tau_m J\chi(f)\mathcal{D}(f), \tag{A.33}$$

which summarizes the average linear response of a neuron's firing rate to a single excitatory spike train. With this notation, eq. (A.32) becomes

$$S^{EE}_{x_1x_2,1\to2} \approx S^{EE}_{x_1x_2,nc} + A^* S_{xx}. \tag{A.34}$$

If the connection is reversed, the indices in eq. (A.31) must be swapped and the cross-spectrum is the complex conjugate of eq. (A.34)

$$S^{EE}_{x_1x_2,1\leftarrow2} \approx S^{EE}_{x_1x_2,nc} + A S_{xx}. \tag{A.35}$$

If the two neurons are reciprocally connected, one obtains

$$\begin{aligned} S^{EE}_{x_1x_2,1\leftrightarrow2} &\approx S^{EE}_{x_1x_2,nc}(1+|A|^2) + 2\Re[A]S_{xx} \\ &\approx S^{EE}_{x_1x_2,nc} + 2\Re[A]S_{xx} \end{aligned} \tag{A.36}$$

where the term $|A|^2 \ll 1$ was neglected.

In the case of an excitatory-inhibitory pair, similar calculations yield

$$S^{EI}_{x_1x_2,1\to 2} \approx S^{EI}_{x_1x_2,nc} + A^* S_{xx} \tag{A.37}$$
$$S^{EI}_{x_1x_2,1\leftarrow 2} \approx S^{EI}_{x_1x_2,nc} - gAS_{xx} \tag{A.38}$$
$$S^{EI}_{x_1x_2,1\leftrightarrow 2} \approx S^{EE}_{x_1x_2,nc} + (A^* - gA)S_{xx},. \tag{A.39}$$

The case of inhibitory-inhibitory pair presents no conceptual difference to the other two. The result is

$$S^{II}_{x_1x_2,1\to 2} \approx S^{II}_{x_1x_2,nc} - gA^* S_{xx} \tag{A.40}$$
$$S^{II}_{x_1x_2,1\leftarrow 2} \approx S^{II}_{x_1x_2,nc} - gAS_{xx} \tag{A.41}$$
$$S^{II}_{x_1x_2,1\leftrightarrow 2} \approx S^{II}_{x_1x_2,nc} - 2g\Re[A]S_{xx},, \tag{A.42}$$

The above linear approximations to the effect of the different direct connection motifs will be now compared to network simulations. Given the large number of possible cases, only the excitatory-inhibitory pair will be shown (all other cases are qualitatively similar). It is convenient to plot the difference between each term on the left side of the three equations and $S^{EI}_{x_1x_2,nc}$. In fig. A.4A the real part of $S^{EI}_{x_1x_2,1\to 2} - S^{EI}_{x_1x_2,nc}$ measured from network simulations is plotted as a red continuous line. The linear approximation according to eq. (A.37), which is $A^* S_{xx}$, is plotted as a red dashed line when the fully theoretical χ is used and as a red dashed-dotted line when $\hat{\chi}$, the susceptibility corrected with the measured firing rate, is used. The same color scheme is used in fig. A.4B to plot the imaginary part of the same quantities. For both the real and imaginary parts, choosing χ or $\hat{\chi}$ does not have a large impact on the linear-response approximation, which captures the effect of a direct excitatory connection rather well, although using $\hat{\chi}$ leads to a slight underestimation. Consider now the case of a direct inhibitory connection from the second neuron to the first. The green line in fig. A.4A (B) shows the real (imaginary) part of $S^{EI}_{x_1x_2,1\leftarrow 2} - S^{EI}_{x_1x_2,nc}$ as measured from network simulations. The linear approximation in eq. (A.38) using the susceptibility χ is plotted as a green dashed line and overestimates (in absolute value) the difference from the non-connected case measured in the simulation results. In this case, the main source of discrepancy is the error in the theoretical spontaneous firing rate: indeed, the agreement with the linear-response theory improves a lot if $\hat{\chi}$ is used (green dashed-dotted line). Analogous considerations apply to the third case considered in fig. A.4, namely $S^{EI}_{x_1x_2,1\leftrightarrow 2} - S^{EI}_{x_1x_2,nc}$. Analogously to the previous cases, blue continuous, dashed, and dotted lines show the numerical simulations, the χ-based, and the $\hat{\chi}$-based linear-response theory, respectively.

To determine what the *average* effect of direct connections is, the four possible connection motifs must be combined while taking the respective probabilities into account. The network

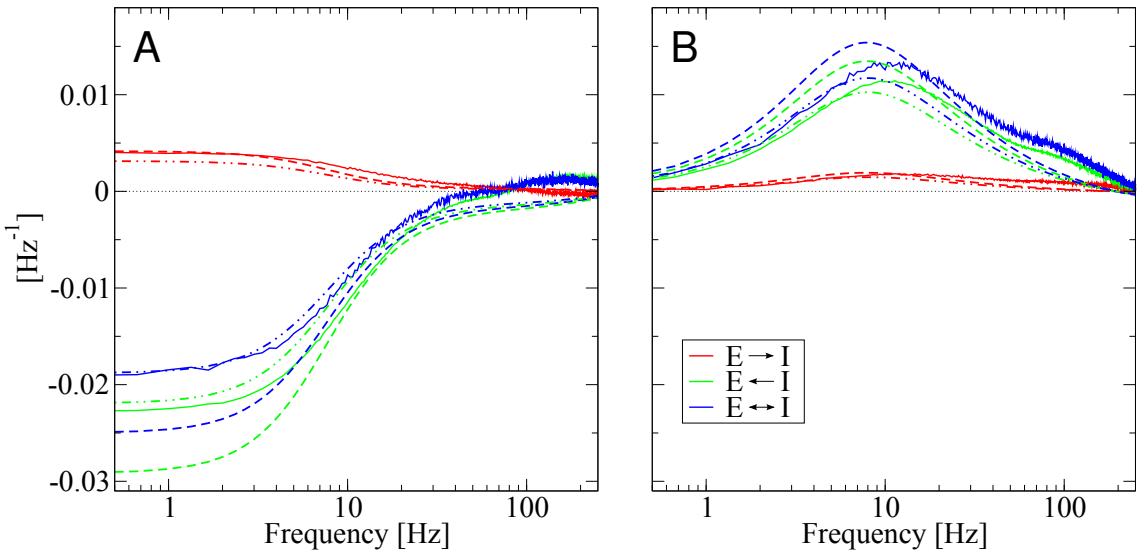

Figure A.4. – **Linear response approximation for the effect of single direct connections on the cross spectrum between an excitatory-inhibitory neuron pair. A**: real part. **B**: imaginary part. Red continuous line: $S^{EI}_{x_1x_2,1\to 2} - S^{EI}_{x_1x_2,nc}$ measured from network simulations. Green continuous line: $S^{EI}_{x_1x_2,1\leftarrow 2} - S^{EI}_{x_1x_2,nc}$ measured from network simulations. Blue continuous line $S^{EI}_{x_1x_2,1\leftrightarrow 2} - S^{EI}_{x_1x_2,nc}$ measured from network simulations. Dashed lines are the linear response theory according to eqs. (A.37) to (A.39) with the same color coding. Dashed-dotted lines are the linear-response-theory-corrected χ_{num} and the measured S_{xx}. Parameters as in table 2.2. The agreement of eqs. (A.37) to (A.39) for the other parameter sets is qualitatively similar.

topology is such that any two randomly selected neurons are connected (in one direction) with probability $p_c \approx C_E/N_E$. Therefore, a given pair of neurons is not directly connected with probability $(1-p_c)^2$, is unidirectionally connected with probability $p_c(1-p_c)$ and reciprocally connected with probability p_c^2. Considering first the average spectrum between two excitatory neurons, $S^{EE}_{x_1x_2}$ can be split according to the four possible cases:

$$S^{EE}_{x_1x_2} = (1-p_c)^2 S^{EE}_{x_1x_2,nc} + p_c(1-p_c) S^{EE}_{x_1x_2,1\to 2} + p_c(1-p_c) S^{EE}_{x_1x_2,1\leftarrow 2} + p_c^2 S^{EE}_{x_1x_2,1\leftrightarrow 2}, \quad \text{(A.43)}$$

remembering that $S^{EE}_{x_1x_2,nc}$ is the average cross-spectrum between two excitatory neurons that are not directly connected to each other, $S^{EE}_{x_1x_2,1\to 2}$ is the average spectrum of an excitatory pair in which the first neuron is connected to the second one (but not vice-versa), $S^{EE}_{x_1x_2,1\leftarrow 2}$ is the opposite case, and $S^{EE}_{x_1x_2,1\leftrightarrow 2}$ is the average cross-spectrum of bidirectionally coupled excitatory neurons. Inserting eqs. (A.34) to (A.36) into eq. (A.43) yields

$$S^{EE}_{x_1x_2} \approx S^{EE}_{x_1x_2,nc} + 2\Re[A] p_c S_{xx}. \quad \text{(A.44)}$$

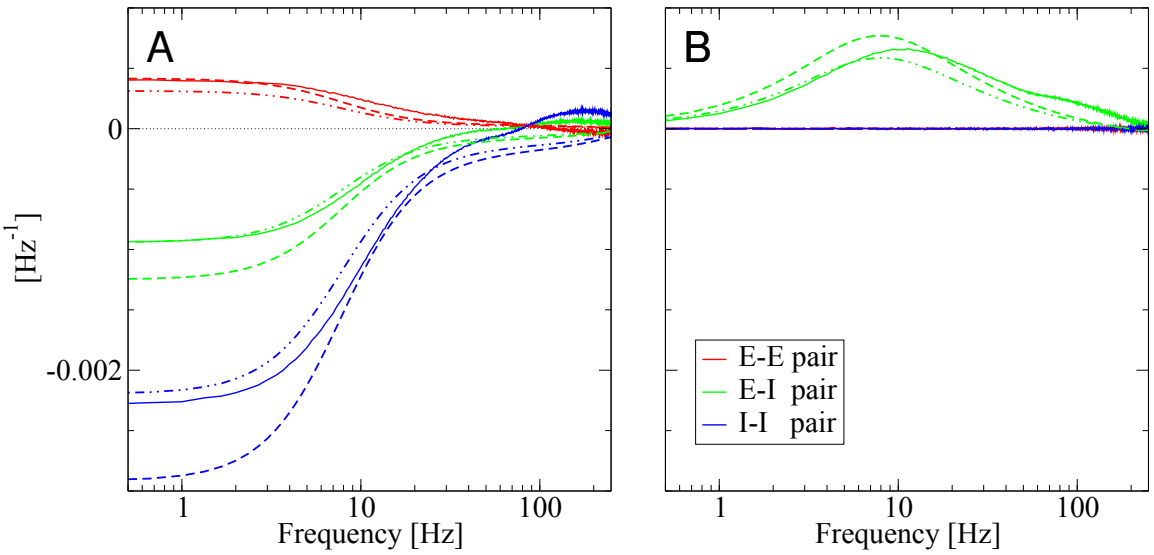

Figure A.5. – **Linear response approximation for the *average* effect of direct connections on the cross spectrum between neuron pairs in the recurrent network. A**: real part. **B**: imaginary part. Red continuous line: $S^{EE}_{x_1x_2} - S^{EE}_{x_1x_2,nc}$ measured from network simulations. Green continuous line: $S^{EI}_{x_1x_2} - S^{EI}_{x_1x_2,nc}$ measured from network simulations. Blue continuous line $S^{II}_{x_1x_2} - S^{II}_{x_1x_2,nc}$ measured from network simulations. With the same meaning of colors, dashed lines represent the linear response theory according to eqs. (A.44) to (A.46). Dashed-dotted lines show theory predictions corrected by using χ_{num} and the measured S_{xx}. Parameters as in table 2.2. The agreement of the theory for the other parameter sets is similar.

The decomposition eq. (A.43) holds also for the other two cases, $S^{EI}_{x_1x_2}$ and $S^{II}_{x_1x_2}$. Combining eqs. (A.37) to (A.39) with an analogous form of eq. (A.43) for $S^{EI}_{x_1x_2}$ leads to

$$S^{EI}_{x_1x_2} \approx S^{EI}_{x_1x_2,nc} + (A^* - gA)p_c S_{xx}. \tag{A.45}$$

Similarly, by combining eqs. (A.40) to (A.42) with the respective probabilities as in eq. (A.43) one finds

$$S^{II}_{x_1x_2} \approx S^{II}_{x_1x_2,nc} - 2g\Re[A]p_c S_{xx}. \tag{A.46}$$

The last three results, eqs. (A.44) to (A.46), can be compared to network simulations in the same fashion as in the previous fig. A.4, i.e. by plotting the difference $S^{XY}_{x_1x_2} - S^{XY}_{x_1x_2,nc}$, where $X, Y = E, I$. This comparison is shown in fig. A.5, in which the meaning of all line styles is the same as in the previous figure: solid lines are network simulations, dashed lines are the theoretical approximations based on χ, and dashed-dotted lines are theoretical approximations based on $\hat{\chi}$. Furthermore, all real parts are shown in fig. A.5A and all imaginary parts in fig. A.5B. The difference $S^{EE}_{x_1x_2} - S^{EE}_{x_1x_2,nc}$ is plotted in red, the difference $S^{EI}_{x_1x_2} - S^{EI}_{x_1x_2,nc}$ is plotted in green, and

the difference $S^{II}_{x_1x_2} - S^{II}_{x_1x_2,nc}$ is plotted in blue. Not surprisingly, the general picture is similar to the previous case: the approximation is better for the EE case (plotted in red), while at low frequencies the effect of inhibitory connections is overestimated by the linear-response theory (dashed lines) which leads to discrepancies for the EI case (green) and the II case (blue). Using the $\hat{\chi}$ approximation (dotted lines) grants a generally better agreement. Intuitively, the effect of direct excitatory connections is to increase the spike count correlation over large time windows (related to the limit $f \to 0$ of the cross-spectrum), and the effect of inhibitory connections is to anti-correlate the spike count over large windows. Interestingly, for frequencies above $\approx 80\,\mathrm{Hz}$, the effect of direct inhibitory connections has the opposite sign, which is completely missing in the linear theory.

From the "conjugate symmetry" of cross-spectra, $S^{XY}_{x_1x_2} = (S^{YX}_{x_1x_2})^*$, it follows that $S^{EI}_{x_1x_2}$ is the only spectrum to have a non-zero imaginary part. In fact, $S^{EE}_{x_1x_2} = (S^{EE}_{x_1x_2})^*$ and $S^{II}_{x_1x_2} = (S^{II}_{x_1x_2})^*$ implies that$S^{EE}_{x_1x_2}$ and $S^{II}_{x_1x_2}$ have no imaginary part, as seen in fig. A.5B.

Impact of paths of length two

Spikes fired from one neuron can influence the second neuron through longer paths. The effect of paths of length two between a excitatory-excitatory neuron pair will be considered first. The first neuron has l_E (l_I) target neurons that are excitatory (inhibitory) and directly connected to the second neuron. In other words, l_E (l_I) is the number of paths of length two connecting the first neuron to the second one, in which the intermediate neuron is excitatory (inhibitory). On average, there are $p_c C_E$ ($p_c \gamma C_E$) such paths. To estimate the effect of paths from neuron 1 to neuron 2, consider the input to neuron 2:

$$\begin{aligned}\eta_2 = \tau_m &\left(\sum_i^{l_E} J_{2p_{21,i}} x^E_{p_{21,i}} e^{2\pi i f D_{2p_{21,i}}} - g \sum_j^{l_I} J_{2q_{21,j}} x^I_{q_{21,j}} e^{2\pi i f D_{2q_{21,j}}} \right) \\ &+ \tau_m \left(\sum_i^{C_E - l_E} J_{2r_{2,i}} x^E_{r_{2,i}} e^{2\pi i f D_{2r_{2,i}}} - g \sum_j^{C_I - l_I} J_{2s_{2,j}} x^I_{s_{2,j}} e^{2\pi i f D_{2s_{2,j}}} \right).\end{aligned} \tag{A.47}$$

In the last line, $r_{2,i}$ ($s_{2,j}$) runs over all excitatory (inhibitory) inputs of neuron two. The first $C_E - l_E$ ($C_I - l_I$) indexes $\{r_{2,i}\}_{i=1\ldots C_E - l_E}$ ($\{s_{2,j}\}_{j=1\ldots C_I - l_I}$) refer to neurons that do *not* receive input from neuron one; the last l_E (l_I) indexes $\{r_{2,i}\}_{i=C_E - l_E + 1\ldots C_E}$ ($\{s_{2,j}\}_{j=C_I - l_I + 1\ldots C_I}$) label presynaptic neurons of neuron two that also receive input from neuron one. In other words, these neurons are those acting as joints for paths of length between neuron 1 and neuron 2. These intermediate neurons are also labeled with $p\{_{21,i}\}_{i=1\ldots l_E}$ ($\{q_{21,j}\}_{j=1\ldots l_I}$), i.e. $r_{2,i} = p_{21,i-l_E}$ for $i > l_E$ and $s_{2,j} = q_{21,j-l_I}$ for $j > l_I$ runs over all excitatory (inhibitory) presynaptic neurons of 2 that are postsynaptic targets of neuron 1. For each of these l_E (l_I) "relay" neurons the

linear-response reads

$$\begin{aligned} x^E_{p_{21,i}} &= x^E_{0,p_{21,i}} + \chi\tau_m J_{p_{21,i}1} e^{2\pi i f D_{p_{21,i}1}} x^E_1 \\ x^I_{q_{21,j}} &= x^I_{0,q_{21,j}} + \chi\tau_m J_{q_{21,j}1} e^{2\pi i f D_{q_{21,j}1}} x^E_1, \end{aligned} \tag{A.48}$$

where $x^E_{0,p_{21,i}}$ ($x^I_{0,q_{21,j}}$) is the spiking activity of neuron $p_{21,i}$ ($q_{21,j}$) if the input connection from neuron 1 to $p_{21,i}$ ($q_{21,j}$) is removed. Inserting the last two equations into eq. (A.47) yields:

$$\begin{aligned} \eta_2 = \tau_m \Bigg(& \sum_i^{l_E} J_{2p_{21,i}} e^{2\pi i f D_{2p_{21,i}}} x^E_{0,p_{21,i}} + \chi\tau_m J_{p_{21,i}1} J_{2p_{21,i}} e^{2\pi i f D_{p_{21,i}1}} x^E_1 e^{2\pi i f D_{2p_{21,i}}} \\ & - g \sum_j^{l_I} J_{2q_{21,j}} e^{2\pi i f D_{2q_{21,j}}} x^I_{0,q_{21,j}} + \chi\tau_m J_{q_{21,j}1} J_{2q_{21,j}} e^{2\pi i f D_{q_{21,j}1}} x^E_1 e^{2\pi i f D_{2q_{21,j}}} \Bigg) \\ & + \tau_m \left(\sum_i^{C_E - l_E} J_{2r_{2,i}} x^E_{r_{2,i}} e^{2\pi i f D_{2r_{2,i}}} - g \sum_j^{C_I - l_I} J_{2s_{2,j}} x^I_{s_{2,j}} e^{2\pi i f D_{2s_{2,j}}} \right) \\ = \tau_m^2 \chi x^E_1 & \left(\sum_i^{l_E} J_{p_{21,i}1} J_{2p_{21,i}} e^{2\pi i f D_{p_{21,i}1}} e^{2\pi i f D_{2p_{21,i}}} - g \sum_j^{l_I} J_{q_{21,j}1} J_{2q_{21,j}} e^{2\pi i f D_{q_{21,j}1}} e^{2\pi i f D_{2q_{21,j}}} \right) \\ & + \tau_m \left(\sum_i^{C_E} J_{2r_{2,i}} x^E_{r_{2,i}} e^{2\pi i f D_{2r_{2,i}}} - g \sum_j^{C_I} J_{2s_{2,j}} x^I_{s_{2,j}} e^{2\pi i f D_{2s_{2,j}}} \right) \\ = \chi^{-1} \alpha_{E,1} x^E_1 & + \eta_{2,0}, \end{aligned} \tag{A.49}$$

where $\eta_{2,0}$ is the input to neuron 2 if the connections from 1 to $\{p_{21,i}\}_{i=1\ldots l_E}$ and $\{q_{21,j}\}_{j=1\ldots l_I}$ are removed. The shorthand

$$\alpha_{E,1} = \tau_m^2 \chi^2 \left(\sum_i^{l_E} J_{p_{21,i}1} J_{2p_{21,i}} e^{2\pi i f D_{p_{21,i}1}} e^{2\pi i f D_{2p_{21,i}}} - g \sum_j^{l_I} J_{q_{21,j}1} J_{2q_{21,j}} e^{2\pi i f D_{q_{21,j}1}} e^{2\pi i f D_{2q_{21,j}}} \right), \tag{A.50}$$

summarizes the linear response of neuron 2 to the summed input incoming from neuron 1 via all paths of length two:

$$x_2 = x_{2,0} + \chi\chi^{-1}\alpha_{E,1} x^E_1 = x_{2,0} + \alpha_{E,1} x^E_1. \tag{A.51}$$

Analogously, the linear response of neuron 1 to all spikes originating from neuron 2 and traveling over paths of length two is

$$x_1 = x_{1,0} + \chi\chi^{-1}\alpha_{E,2} x^E_2 = x_{1,0} + \alpha_{E,2} x^E_2, \tag{A.52}$$

where $x_{1,0}$ is defined as the spiking activity of neuron 1 if spikes fired from neuron 2 to the intermediary neurons are removed. Solving for x_1, x_2 yields:

$$\begin{aligned} x_1 &= \frac{x_{1,0} + \alpha_{E,2} x_{2,0}}{1 - \alpha_{E,1}\alpha_{E,2}} \approx x_{1,0} + \alpha_{E,2} x_{2,0} \\ x_2 &= \frac{x_{2,0} + \alpha_{E,1} x_{1,0}}{1 - \alpha_{E,1}\alpha_{E,2}} \approx x_{2,0} + \alpha_{E,1} x_{1,0}, \end{aligned} \tag{A.53}$$

where $\alpha_{E,1}\alpha_{E,2} \ll 1$ even in a very unlikely realization of the two sums in eq. (A.50), in which for instance all l_E excitatory terms are ten times as large as the average and all inhibitory terms are much smaller than average, or the other way around. As they consist of a fairly large ($l_E = 200$ and $l_I = 50$ for the SAN and SDN) number of i.i.d. terms, $\alpha_{E,1}$ and $\alpha_{E,2}$ are more likely to be relatively close to the mean value, i.e. $\alpha_{E,1}/\langle\alpha_{E,1}\rangle \approx 1$, where

$$\alpha_E = \langle\alpha_{E,1}\rangle = \langle\alpha_{E,2}\rangle = \tau_m^2 \chi^2 J^2 \left(\mathcal{D}(f)\right)^2 \langle l_E - g l_I\rangle = A^2 p_c^2 N_E (1 - g\gamma). \tag{A.54}$$

The term $x_{1,0}$ represents the spiking activity of neuron 1 if spikes from neuron 2 to neuron 1 are not considered; the effect of these spikes is described by $\alpha_{E,2}$. Hence, $x_{1,0}$ and $\alpha_{E,2}$ are uncorrelated by definition. Furthermore, because $\alpha_{E,1}$ summarizes the effect of paths *originating* from neuron 1 terminating on neuron 2, it is reasonable to assume (as discussed in section A.3) that $x_{1,0}$ is uncorrelated with $\alpha_{E,1}$ as well. Since the labels are arbitrary, $x_{2,0}$ is also uncorrelated with $\alpha_{E,1}$ and $\alpha_{E,2}$. From these considerations and from eq. (A.53) it follows that the cross-spectrum reads:

$$\begin{aligned} S^{EE}_{x_1x_2,nc} &= \frac{1}{T}\langle x_1 x_2^*\rangle = S^{FF}_{x_1x_2}\left\langle 1 + \alpha^*_{E,1}\alpha_{E,2}\right\rangle + \left\langle \alpha^*_{E,1} + \alpha_{E,2}\right\rangle S_{xx} + \dots \\ &\approx S^{FF}_{x_1x_2} + (\langle\alpha^*_{E,1}\rangle + \langle\alpha_{E,2}\rangle) S_{xx} + \dots = S^{FF}_{x_1x_2} + 2\Re[\alpha_E] S_{xx} + \dots \end{aligned} \tag{A.55}$$

where the dots stand for the effect of loops of length larger than two, which have been neglected, and the fact that $S^{FF}_{x_1x_2} = (S^{FF}_{x_1x_2})^*$ was used.

If neuron two is inhibitory, its effect on neuron one only differs from eq. (A.52) by a factor $-g$:

$$x_1 = x_{1,0} + \chi\chi^{-1}\alpha_{I,2} x_2^E = x_{1,0} - g\alpha_{E,2} x_2^E. \tag{A.56}$$

By carrying out the same calculation as above, one obtains

$$\begin{aligned} S^{EI}_{x_1x_2,nc} &= \frac{1}{T}\langle x_1 x_2^*\rangle = S^{FF}_{x_1x_2}\left\langle 1 - g\alpha^*_{E,1}\alpha_{E,2}\right\rangle + \left\langle \alpha^*_{E,1} - g\alpha_{E,2}\right\rangle S_{xx} + \dots \\ &\approx S^{FF}_{x_1x_2} + (\langle\alpha^*_{E,1}\rangle - g\langle\alpha_{E,2}\rangle) S_{xx} + \dots = S^{FF}_{x_1x_2} + (\alpha^*_E - g\alpha_E) S_{xx}. \end{aligned} \tag{A.57}$$

If both neurons are inhibitory, the linear response looks like eq. (A.56) for both neurons, so that

the resulting cross-spectrum is

$$\begin{aligned} S^{II}_{x_1x_2,nc} = \frac{1}{T}\langle x_1 x_2^* \rangle &= S^{FF}_{x_1x_2}\left\langle 1 + g^2 \alpha^*_{E,1}\alpha_{E,2} \right\rangle - g\left\langle \alpha^*_{E,1} + \alpha_{E,2} \right\rangle S_{xx} + \dots \\ &\approx S^{FF}_{x_1x_2} - g(\langle \alpha^*_{E,1}\rangle + \langle \alpha_{E,2}\rangle) S_{xx} + \dots = S^{FF}_{x_1x_2} - 2g\Re[\alpha_E] S_{xx} + \dots \end{aligned} \tag{A.58}$$

Impact of paths of arbitrary length

The linear-response eq. (A.51) can now be generalized to consider the effect of paths of arbitrary length on a couple of neurons that are not directly connected. More precisely, an expansion (here $X = E, I$ indicates the type of the first neuron) is sought like the following

$$x_2 \approx x_{2,0} + \sum_{\ell=2}^{\infty} \mathcal{L}^X_{\ell,1} x_1^X = x_{2,0} + \mathcal{L}^X_{\infty,1} x_1^X \tag{A.59}$$

where $x_{2,0}$ is the activity of neuron two if the effect of the spikes fired from neuron one is removed from the network, and $\sum_{\ell=2}^{\infty} \mathcal{L}^X_{\ell,1} = \mathcal{L}^X_{\infty,1}$ summarizes the effect of spikes fired by neuron one reaching neuron two via paths of length $\ell > 1$. Note that this ansatz is similar in spirit but different from that by Trousdale et al. (2012), in that here $x_{2,0}$ represents the activity of neuron two *when only spikes from neuron one* are removed, whereas Trousdale et al. (2012) consider as unperturbed state the activity of neurons driven only by the external noise, i.e. if the *entire* network input is ignored. Assuming that $\mathcal{L}^X_{\infty,1} \ll 1$ by analogy with eq. (A.53) leads to ($Y = E, I$ indicates the type of neuron two)

$$\begin{aligned} x_1 &= \frac{x_{1,0} + \mathcal{L}^Y_{\infty,2} x_{2,0}}{1 - \mathcal{L}^X_{\infty,1}\mathcal{L}^Y_{\infty,2}} \approx x_{1,0} + \mathcal{L}^Y_{\infty,2} x_{2,0} \\ x_2 &= \frac{x_{2,0} + \mathcal{L}^X_{\infty,1} x_{1,0}}{1 - \mathcal{L}^X_{\infty,1}\mathcal{L}^Y_{\infty,2}} \approx x_{2,0} + \mathcal{L}^X_{\infty,1} x_{1,0}. \end{aligned} \tag{A.60}$$

By virtue of the same argument used for $\alpha_{E,1}$ and $\alpha_{E,2}$ in the last subsection, it can be assumed that $\mathcal{L}^X_{\infty,1}$ and $\mathcal{L}^Y_{\infty,2}$ are uncorrelated with $x_{1,0}$ and $x_{2,0}$. Therefore, inserting eq. (A.60) into the definition of the cross-spectrum between x_1 and x_2 yields

$$S^{XY}_{x_1x_2,nc} = \frac{1}{T}\langle x_1 x_2^* \rangle \approx S^{FF}_{x_1x_2} + [\langle (\mathcal{L}^X_{\infty,1})^* \rangle + \langle \mathcal{L}^Y_{\infty,2} \rangle] S_{xx} = S^{FF}_{x_1x_2} + [(\mathcal{L}^X_{\infty})^* + \mathcal{L}^Y_{\infty}] S_{xx}, \tag{A.61}$$

where $\mathcal{L}^X_{\infty} = \langle \mathcal{L}^X_{\infty,1} \rangle = \langle \mathcal{L}^X_{\infty,2} \rangle$ and $\mathcal{L}^X_{\ell} = \langle \mathcal{L}^X_{\ell,1} \rangle = \langle \mathcal{L}^X_{\ell,2} \rangle$.

It is known from the last subsection that $\mathcal{L}^E_{\ell=2} = \alpha_E$, and that $\mathcal{L}^I_{\ell=2} = -g\alpha_E$. As a preliminary step to the calculation of the other terms, the result $\mathcal{L}^E_2 = \alpha_E$ will be derived again in a more synthetic way. The situation is portrayed in the top row of fig. A.6, in which Ω_1 is the entire recurrent network and only connections from neuron one and to neuron two are schematically

depicted. *On average*, the effect of the spike train x_1 onto neuron two via *a single* path of length two is either $A^2 x_1$, if it travels via an intermediate excitatory neuron and thus through two excitatory synapses, or $-gA^2 x_1$, if the intermediate neuron is inhibitory and therefore spikes first go through one excitatory synapse and then through an inhibitory synapse. How many paths of length two exist? For a path of length two to form, the N_E excitatory neurons in Ω_1 must receive input from neuron one and send output to neuron two. Because both input and output connections are present with probability $\approx p_c$, there are on average $p_c^2 N_E$ paths with an intermediary excitatory neuron and $p_c^2 \gamma N_E$ paths via one inhibitory neuron. Summing up the contribution of the two types of paths yields

$$\mathcal{L}_2^E = \langle \alpha_{EE,1} \rangle = A^2 p_c^2 N_E (1 - g\gamma), \tag{A.62}$$

which is the result derived before. If x_1 is inhibitory, the only difference is that the first outgoing synapse is inhibitory, so that $\mathcal{L}_2^I$ reads

$$\mathcal{L}_2^I = -g\mathcal{L}_2^E, \tag{A.63}$$

where the factor $-g$ accounts for the first inhibitory output synapse.

Consider now a path of length three, i.e. spikes traveling via two intermediate neurons before reaching neuron two, as depicted in the middle row of fig. A.6, where Ω_2 and Ω_1 represent two "virtual" copies of the same recurrent network. If the same approach as in the previous case is used, the four possibilities to travel from neuron one to neuron two must be distinguished according to the type of the two intermediate neurons. Each path contributes to $\mathcal{L}_3$ with a term $(-g)^i A^3 x_1$, where $i = 0, 1, 2$ is the number of inhibitory neurons present in the considered path. There are $n_{EE} = p_c^3 N_E^2$ paths via two excitatory neurons, $n_{EI} = n_{IE} = p_c^3 N_E^2 \gamma$ paths passing through one inhibitory and one excitatory neuron, and $n_{II} = p_c^3 N_E^2 \gamma^2$ paths via two inhibitory neurons. Therefore,

$$\mathcal{L}_3^E = (n_{EE} - g(n_{EI} + n_{IE}) + n_{II} g^2) A^3 = A^3 p_c^3 N_E^2 (1 - g\gamma)^2. \tag{A.64}$$

This last result can be obtained by an alternative "recursive" approach that is more amenable to generalization. Each excitatory neuron within Ω_2 is connected by paths of length two to neuron two, and the combined effect of all these paths is $\mathcal{L}_2^E$. At the same time, there are $p_c N_E$ neurons in Ω_2 that also *receive* input from neuron one; these neurons are therefore the first joint for all paths of length *three* connecting neuron one to neuron two, and each of them yields a contribution $A \cdot \mathcal{L}_2$. Analogously, all $p_c \gamma N_I$ inhibitory neurons within Ω_2 that are receiving input from neuron one are also connected to neuron two via paths of length two that give a total

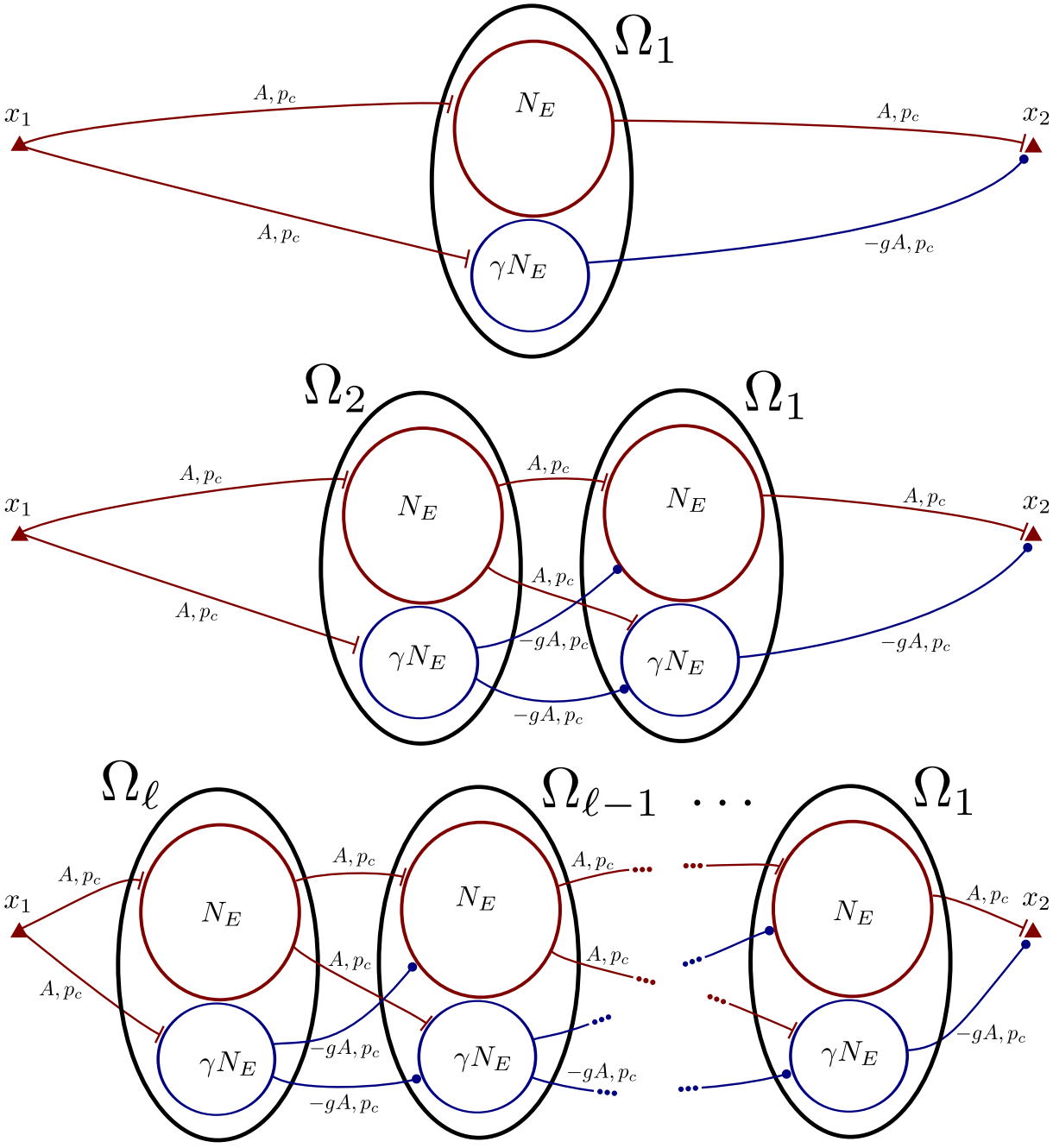

Figure A.6. – Visualization of the linear-order contribution of paths of arbitrary length to cross-correlations between two excitatory neurons. Constructing multiple copies of the network (Ω_ℓ) is a useful device to determine the number of possible paths and their total contribution to cross-correlations. As explained in the main text, if the first neuron is inhibitory, the total contribution of paths of length ℓ is multiplied by a factor $-g$.

average contribution of $\mathcal{L}_2^I$. Therefore, the summed effect of paths of length three is

$$\mathcal{L}_3^E = p_c N_E A \mathcal{L}_2^E + p_c \gamma N_E A \mathcal{L}_2^I = A p_c N_E (1 - g\gamma) \mathcal{L}_2^E = A^3 p_c^3 N_E^2 (1 - g\gamma)^2, \tag{A.65}$$

which is consistent with the previous result. The effect of paths of length three originating from an inhibitory neuron is again simply $\mathcal{L}_3^I = -g\mathcal{L}_3^E$.

It is easy to generalize the last argument to paths of arbitrary length ℓ. Consider the bottom row in fig. A.6: neuron 1 connects to $p_c N_E$ excitatory neurons and to $p_c \gamma N_E$ inhibitory neurons within Ω_ℓ; the former affect neuron 2 with a term $\mathcal{L}_{\ell-1}^E$, the latter affect neuron 2 with a term $\mathcal{L}_{\ell-1}^I$. Therefore, the total effect of paths of length ℓ is

$$\mathcal{L}_\ell^E = p_c N_E A \mathcal{L}_{\ell-1}^E + p_c \gamma N_E A \mathcal{L}_{\ell-1}^I = A p_c N_E (1 - g\gamma) \mathcal{L}_{\ell-1}^E = A^\ell p_c^\ell N_E^{\ell-1} (1 - g\gamma)^{\ell-1}, \tag{A.66}$$

where in the last equality the "inductive hypothesis" derived by the previous cases $\ell = 2$ and $\ell = 3$ was inserted. As usual, if the source neuron is inhibitory, the first outgoing inhibitory synapse causes a factor $-g$ to appear

$$\mathcal{L}_\ell^I = -g\mathcal{L}_\ell^E. \tag{A.67}$$

Assuming convergence of the series, the combined effect of paths of all lengths is

$$\mathcal{L}_\infty^E = \sum_{\ell=2}^{\infty} A^\ell p_c^\ell N_E^{\ell-1} (1 - g\gamma)^{\ell-1} = A p_c \sum_{\ell'=1}^{\infty} \left[A p_c N_E (1 - g\gamma) \right]^{\ell'} = A p_c (\beta_\mathcal{L} - 1), \tag{A.68}$$

where

$$\beta_\mathcal{L} = \frac{1}{1 - A p_c N_E (1 - g\gamma)}. \tag{A.69}$$

Using eq. (A.68) and

$$\mathcal{L}_\infty^I = -g\mathcal{L}_\infty^E = -g A p_c (\beta_\mathcal{L} - 1) \tag{A.70}$$

together with eq. (A.61) yields the following expressions:

$$S_{x_1 x_2, nc}^{EE} \approx S_{x_1 x_2}^{FF} + 2 p_c \Re[A(\beta_\mathcal{L} - 1)] S_{xx} \tag{A.71}$$

$$S_{x_1 x_2, nc}^{EI} \approx S_{x_1 x_2}^{FF} + p_c [A^* (\beta_\mathcal{L} - 1)^* - g A (\beta_\mathcal{L} - 1)] S_{xx} \tag{A.72}$$

$$S_{x_1 x_2, nc}^{II} \approx S_{x_1 x_2}^{FF} - 2 g p_c \Re[A(\beta_\mathcal{L} - 1)] S_{xx}, \tag{A.73}$$

which summarize the effect (to linear order) of all possible non-direct paths between the pair of considered neurons. Interestingly, if the network recurrent input is exactly "balanced", i.e. $g = \gamma^{-1}$, then $\beta_\mathcal{L} = 1$ and the effect of all paths cancels out. The same fact was also pointed out by Trousdale et al. (2012).

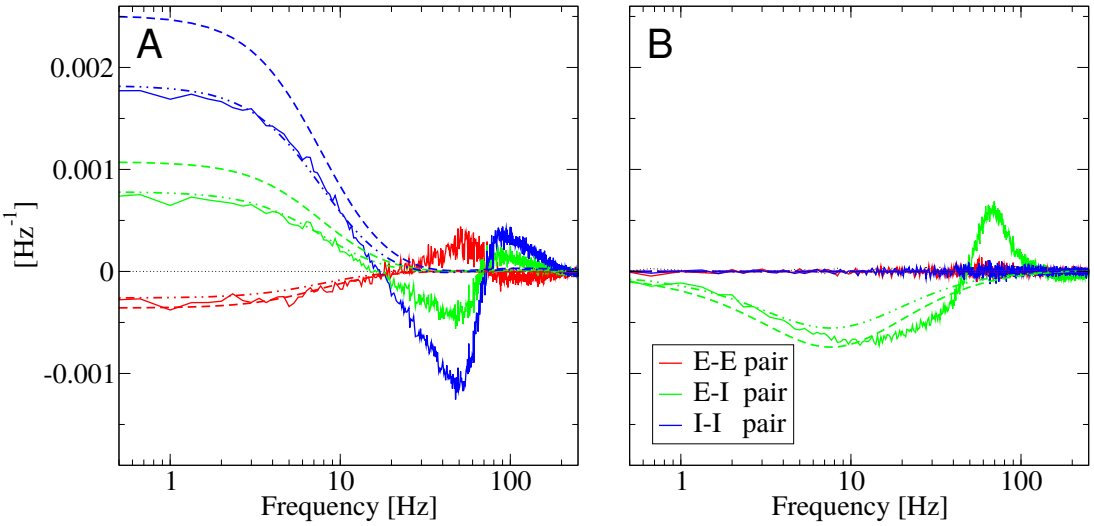

Figure A.7. – **Linear response approximation for the average effect of paths of length $\ell > 1$ on the cross spectrum between neuron pairs with no direct connection. A**: real part of $S^{XY}_{x_1x_2,nc} - S^{FF}_{x_1x_2}$, where $X, Y = E, I$. **B**: imaginary part. The color code is as in the last two figures, i.e. red for excitatory-excitatory pairs, green for excitatory-inhibitory pairs, and blue for inhibitory-inhibitory pairs. Continuous lines are simulation results, dashed lines represent the linear response theory according to eqs. (A.71) to (A.73). Dashed-dotted lines indicate theoretical predictions corrected by using χ_{num} and the measured S_{xx}. Parameters as in table 2.2. The agreement of the theory for the other standard parameter set (without external noise, table 2.1) is very similar. For the "single-barrel" parameter set the difference $S^{XY}_{x_1x_2,nc} - S^{FF}_{x_1x_2}$ is very small and hard to measure despite extensive averaging. The relative error of the theory is large, especially for EI and II pairs. However, for this parameter set the contribution of longer paths to cross-correlations is negligible compared to direct connections and shared input.

The comparison of the last expressions with numerical simulations is shown in fig. A.7 for the SDN parameters. As was done for the case of direct connections, fig. A.7A(B) shows the real (imaginary) part of the three possible cases $S^{XY}_{x_1x_2,nc} - S^{FF}_{x_1x_2}$; the case $X = Y = E$ is plotted in red, $X = E, Y = I$ is plotted in green, and $X = Y = I$ is plotted in blue. As in the previous plots, the largest discrepancy between simulations (continuous lines) and linear response theory is observed for $S^{EI}_{x_1x_2,nc}$ and $S^{II}_{x_1x_2,nc}$ when the susceptibility χ is used to compute the theory (dashed lines). However, if the numerical susceptibility $\hat{\chi}$ is used to compute the theory (dashed-dotted lines), the agreement is fairly good in all cases.

Remarkably, for the SDN parameters used in fig. A.7, the term $Ap_cN_E(1 - g\gamma) \approx -4 < -1$ so that the series $\sum_\ell \mathcal{L}^X_\ell$ does not converge. As already mentioned at the beginning of this appendix, Trousdale et al. (2012) proceed by expanding their initial ansatz using a series. They argue that, in some cases, the series expansion may not formally converge, even if the initial ansatz remains valid. Although here the calculation proceeds the other way around, a similar

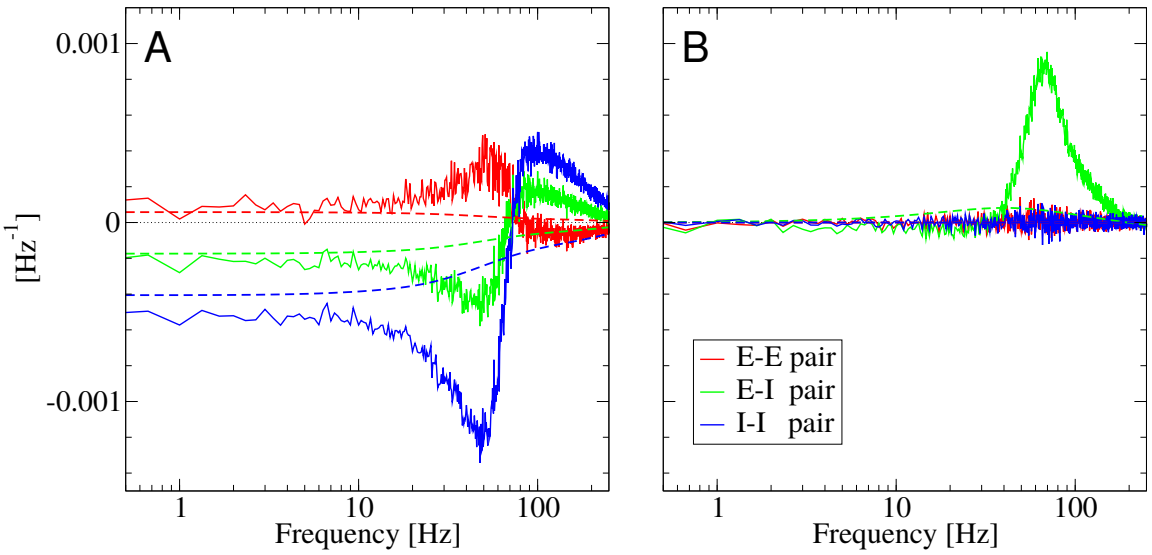

Figure A.8. – **Linear response approximation for the effect of all possible connections on the cross spectrum between neuron pairs in the recurrent network. A**: real part of $S^{XY}_{x_1x_2} - S^{FF}_{x_1x_2}$, where X and Y can be either excitatory or inhibitory. **B**: imaginary part. The meaning of colors code is the same as in the previous figures, i.e. red for excitatory-excitatory pairs, green for excitatory-inhibitory pairs, and blue for inhibitory-inhibitory pairs. Continuous lines represent simulation results and dashed lines stand for the linear response theory according to eqs. (A.74) to (A.76). Parameters as in table 2.2. The agreement of the theory for the parameter set without noise is similar. For the "single-barrel" parameter set the agreement of the theory is better, although a weak oscillation seen at high frequencies is missing.

interpretation can be applied: *if* the series converges, then $\sum_\ell \mathcal{L}^X_\ell = \mathcal{L}^E_\infty$. Otherwise, the decomposition in paths of various lengths does not hold, but the expression for $\mathcal{L}^E_\infty$ may still remain valid.

What happens when $\ell = 1$ in the general expression for $\mathcal{L}^X_\ell$, eqs. (A.66) and (A.67)? The results, $\mathcal{L}^E_1 = Ap_c$ and $\mathcal{L}^I_1 = -gAp_c$, are consistent with the results for direct connections obtained in the previous section eqs. (A.44) to (A.46), as it should be because paths of length $\ell = 1$ are nothing but direct connections. Consequently, combining the previous results eqs. (A.44) to (A.46) with eqs. (A.71) to (A.73) is equivalent to including the case $\ell = 1$ into the sum $\sum_\ell \mathcal{L}^X_\ell$, which yields the rather compact expressions

$$S^{EE}_{x_1x_2} \approx S^{FF}_{x_1x_2} + p_c 2\Re[A\beta_{\mathcal{L}}]S_{xx} \tag{A.74}$$

$$S^{EI}_{x_1x_2} \approx S^{FF}_{x_1x_2} + p_c\left[(A\beta_{\mathcal{L}})^* - gA\beta_{\mathcal{L}}\right]S_{xx} \tag{A.75}$$

$$S^{II}_{x_1x_2} \approx S^{FF}_{x_1x_2} - p_c g 2\Re[A\beta_{\mathcal{L}}]S_{xx}. \tag{A.76}$$

Figure A.8A(B) shows the real (imaginary) part of $S^{XY}_{x_1x_2} - S^{FF}_{x_1x_2}$ measured from network simulations together with the prediction of eqs. (A.74) to (A.76). The convention used for colors and line styles is the same as in the previous plots: red for E-E pairs, green for E-I pairs, and blue for I-I pairs; continuous lines for simulations, dashed lines for linear response prediction. The linear response theory is in reasonable agreement with simulations at low frequencies and is best for E-E and E-I pairs. For higher frequencies, the linear response is completely missing the oscillatory response and provides an inadequate description of the system.

A.5. Putting the pieces together

If eqs. (A.74) to (A.76) are inserted into eqs. (A.5) and (A.29), a rather lengthy equation is obtained, in which only S_{xx} and $S^{FF}_{x_1x_2}$ appear as spectral measures. Solving for $S^{FF}_{x_1x_2}$ yields:

$$S^{FF}_{x_1x_2} \approx |A|^2 p_c S_{xx} \frac{C_E(1+g^2\gamma) + 2Z\Re[A\beta_{\mathcal{L}}]}{1-|A|^2 W C_E^2} \tag{A.77}$$

where

$$Z = C_E^2 \left[(1-g\gamma)(1+g^2\gamma) - \frac{1}{N_E}(1-g^3\gamma)\right] \tag{A.78}$$

and

$$W = (1-g\gamma)^2 - \frac{1+g^2\gamma}{N_E}. \tag{A.79}$$

If the two terms of order $1/N_E$ in Z and W are neglected, eq. (A.77) simplifies to

$$S^{FF}_{x_1x_2} \approx |A|^2 p_c C_E(1+g^2\gamma) \frac{1+2C_E(1-g\gamma)\Re[A\beta_{\mathcal{L}}]}{1-|A(1-g\gamma)C_E|^2} S_{xx}. \tag{A.80}$$

After some algebraic manipulation, the large fraction in the last equation reads as follows,

$$\frac{1+2C_E(1-g\gamma)\Re[A\beta_{\mathcal{L}}]}{1-|A(1-g\gamma)C_E|^2} = \frac{1-C_E(1-g\gamma)|\beta_{\mathcal{L}}|^2(C_E(1-g\gamma)|A|^2-\Re[A^*])}{|\beta_{\mathcal{L}}|^{-2}\{1-C_E(1-g\gamma)|\beta_{\mathcal{L}}|^2(C_E(1-g\gamma)|A|^2-\Re[A])\}} = |\beta_{\mathcal{L}}|^2 \tag{A.81}$$

so that the final result for $S^{FF}_{x_1x_2}$ simplifies to

$$S^{FF}_{x_1x_2} \approx |A|^2 p_c C_E(1+g^2\gamma)|\beta_{\mathcal{L}}|^2 S_{xx} \tag{A.82}$$

Either eq. (A.77) or eq. (A.82) can be used together with eqs. (A.74) to (A.76) to find $S^{EE}_{x_1x_2}$, $S^{EI}_{x_1x_2}$, and $S^{II}_{x_1x_2}$ as functions of the network parameters and of the single-neuron power spectrum. To find the average cross-spectrum between two neurons in the recurrent network, each of the three cross-spectra must be weighted with the respective number of EE, EI, IE, and II pairs, recalling that $S^{IE}_{x_1x_2} = (S^{EI}_{x_1x_2})^*$. Because the network size is large, the approximation

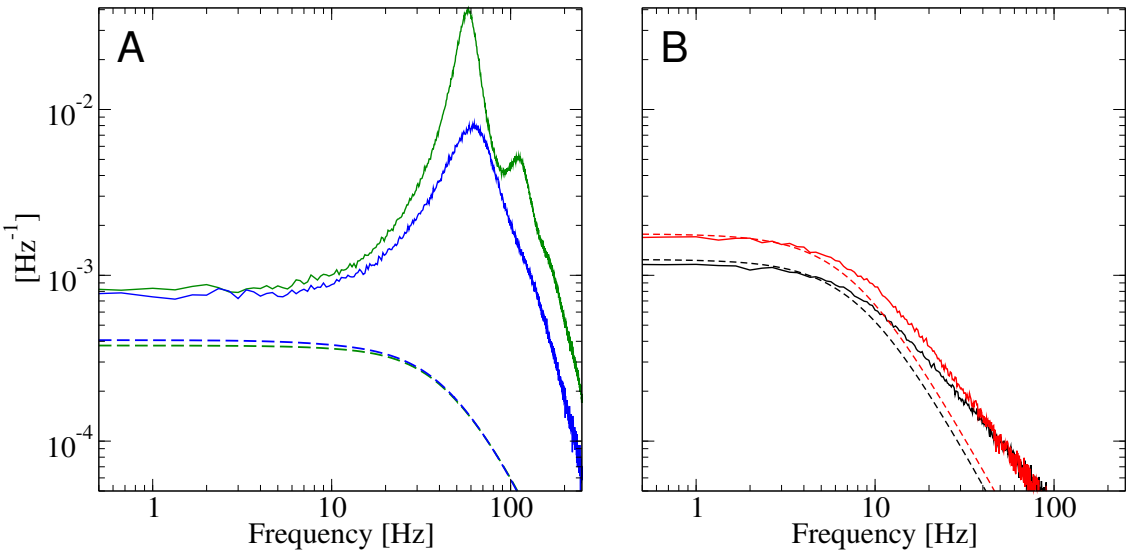

Figure A.9. – Comparison of final result of the linear-response theory eq. (A.83) with network simulations. As the average cross-spectrum between neurons in the network $S_{x_1x_2}$ is real, the imaginary part is not shown. Continuous lines are network simulations and dashed lines are the the linear response theory eq. (A.83). **A**: Standard network parameters in the presence (parameters as in table 2.2, blue lines) and in the absence (parameters as in table 2.1, shown in green) of external shot-noise input. **B**: "Single-barrel" parameters (table 2.3), where the black line represents the inhibition-dominated case ($g = 4.5$) and the red line refers to the balanced case ($g = 4$).

$N(N-1) \approx N^2$ can be made. The final result, the average cross-spectrum between two neurons in the recurrent network, takes the rather compact form:

$$\begin{aligned} S_{x_1x_2} &= \left(\frac{N_E}{N_E(1+\gamma)}\right)^2 S^{EE}_{x_1x_2} + \gamma\left(\frac{N_E}{N_E(1+\gamma)}\right)^2 (S^{EI}_{x_1x_2} + S^{IE}_{x_1x_2}) + \left(\frac{\gamma N_E}{N_E(1+\gamma)}\right)^2 S^{II}_{x_1x_2} \\ &= S^{FF}_{x_1x_2} + \frac{2p_c\Re[A\beta_{\mathcal{L}}]}{1+\gamma}(1-g\gamma)S_{xx} \\ &= S_{xx}p_c\left(|A|^2 C_E(1+g^2\gamma)|\beta_{\mathcal{L}}|^2 + 2\Re[A\beta_{\mathcal{L}}]\frac{1-g\gamma}{1+\gamma}\right) \\ &= S_{xx}p_c\left(|A|^2\frac{C_E(1+g^2\gamma)}{|1-AC_E(1-g\gamma)|^2} + 2\Re\left[\frac{A}{1-AC_E(1-g\gamma)}\right]\frac{1-g\gamma}{1+\gamma}\right), \end{aligned} \tag{A.83}$$

where $A(f) = \tau_m J\chi(f)\mathcal{D}(f)$ was defined in eq. (A.33). For all parameter sets considered, the second term in eq. (A.83) is a rather small correction compared to the first one, so that $S_{x_1x_2} \approx S^{FF}_{x_1x_2}$. A comparison of eq. (A.83) to network simulations is shown in fig. A.9. Results for the SAN and SDN are shown in fig. A.9A in green and blue, respectively. In the range up to about 10 Hz, the theory (dashed line) eq. (A.83) underestimates the cross-spectrum measured

in network simulations (continuous line) by about 50% in both cases. For higher frequencies, the measured cross-spectrum shows one (with external noise) or more (without external noise) pronounced peaks corresponding to global network oscillations. These oscillations are completely missing in the theory, which falls off monotonically. The other panel, fig. A.9B, shows results for the SBN parameter set, which occurs in two variations: "balanced" (red) and "inhibition-dominated" (black) recurrent input. In both cases, the linear response theory (dashed lines) is rather close to the measured cross-spectrum at lower frequencies but decreases faster for increasing frequencies. As already mentioned in chapter 2, the readout filter decreases very rapidly above $f = 10\,\mathrm{Hz}$ so that a fair approximation for the low-frequency limit of $S_{x_1x_2}$ suffices to estimate the variance of the readout activity by using eq. (2.65). Therefore, for the two "standard" parameter sets (fig. A.9A) the theoretical approximation eq. (A.83) is used only to describe the qualitative behavior of cross-correlations as a function of the network size and not to calculate detection rates; for the two "single-barrel" parameter sets (fig. A.9B), eq. (A.83) is used together with the detection theory discussed in section 2.3.3 to obtain a fully analytical estimate of the effect size.

It is worth noting that eq. (A.83) is the same as eq. (25) by Trousdale et al. (2012), which is a bit surprising considering that their initial ansatz, eq. (A.1), does not hold in the absence of external drive. However, thinking of the zero-order term in eq. (A.1) as the activity of the neuron resulting from the "uncorrelated part" of the network input in eq. (A.2) may be a way to reconcile the two approaches. In this interpretation, it is not so striking and rather reassuring that the two calculations prove to be equivalent, because both sum linear-order approximations from all correlation sources.

Trousdale et al. (2012) investigated the discrepancies between actual cross-correlations and the linear-response approximation systematically. They found that the agreement of the linear theory with simulations deteriorates when the recurrent coupling is increased and the external noise is reduced. These observations are in line with the results of this appendix. In fact, the agreement is good for the SBN parameter set and the discrepancies at higher frequencies are possibly due to the crude approximation of the frequency-dependent part of the single neuron susceptibility. Compared to the SBN, the average recurrent coupling strength in the SAN and SDN parameter sets is larger by a factor five for the excitatory coupling and by about a factor eight for the inhibitory coupling (considering both the change in the weight of single synapses and in the number of connections per neuron). Furthermore, the external noise is completely absent in the SAN and much smaller in the SDN (the total excitatory shot noise input rate is reduced by $\approx 75\%$, and the inhibitory shot noise input by 100%). The stronger coupling and weaker external noise are the most likely reasons for the worse agreement of the linear-theory in fig. A.9A. Interestingly, the external shot-noise of the SDN barely influences the cross-correlations in the low-frequency range, whereas it does affect the amplitude and coherence

of the global oscillation at $\approx 60\,\mathrm{Hz}$ (the peak in fig. A.9A), as already noticed in section 2.1.

Appendix B.

Technical Aspects of Detection Procedures

This appendix gives an account of technical advantages and shortcomings of some detection procedures defined in this thesis and in a previous publication (Bernardi and Lindner, 2017). The technical differences between the single-barrier detector (used in chapter 3) and the double-barrier detector (used in chapter 2) are discussed first, followed by a comparison between the definition of effect size used in this thesis and the definition based on the optimal threshold used by Bernardi and Lindner (2017). Importantly, the main findings do not hinge on choosing one of the detection schemes.

B.1. Single-barrier vs. double-barrier detector

The working principle of the double-barrier detector of chapter 2 is explained in section 2.3 (see fig. 2.5). Briefly, one trial can result in a *false positive event* if the input, i.e. the readout activity, is found at least once outside the interval $(R_{\mathrm{sp}} - \theta_d, R_{\mathrm{sp}} + \theta_d)$ during the time window $(-T_w, 0)$. Similarly, a hit (correct detection) is registered if the activity exceeds the upper threshold $R_{\mathrm{sp}} + \theta_d$ or falls below $R_{\mathrm{sp}} - \theta_d$ at least once in $(0, T_w)$. In chapter 3, this detector was replaced by *two* single-barrier detectors. Hits and false positives are defined in the same way as before, but each of the two detectors reacts to crossings of a single barrier: the first detector is equipped with an upper boundary θ_+ and the second one with a lower boundary θ_- (see fig. 3.2). For brevity, the position of the threshold and the corresponding detector are indicated with the same symbol, as in chapter 3. The position of the threshold determines the responsiveness of the detector. To define the *effect size* $\bar{\mathcal{Y}}$, the threshold corresponding to a false positive rate of 0.25 was chosen for each detector. The three thresholds are indicated with $\bar{\theta}_+$, $\bar{\theta}_-$, $\bar{\theta}_d$, respectively. For instance, for the θ_+ detector:

$$\bar{\mathcal{Y}} = \mathcal{Y}(\theta_+) = \mathcal{CD}(\bar{\theta}_+) - \mathcal{FP}(\bar{\theta}_+) = \mathcal{CD}(\bar{\theta}_+) - 0.25, \tag{B.1}$$

where $\mathcal{CD}(\theta_+)$ and $\mathcal{FP}(\theta_+)$ are correct detection and false positive rate as a function of the threshold, respectively.

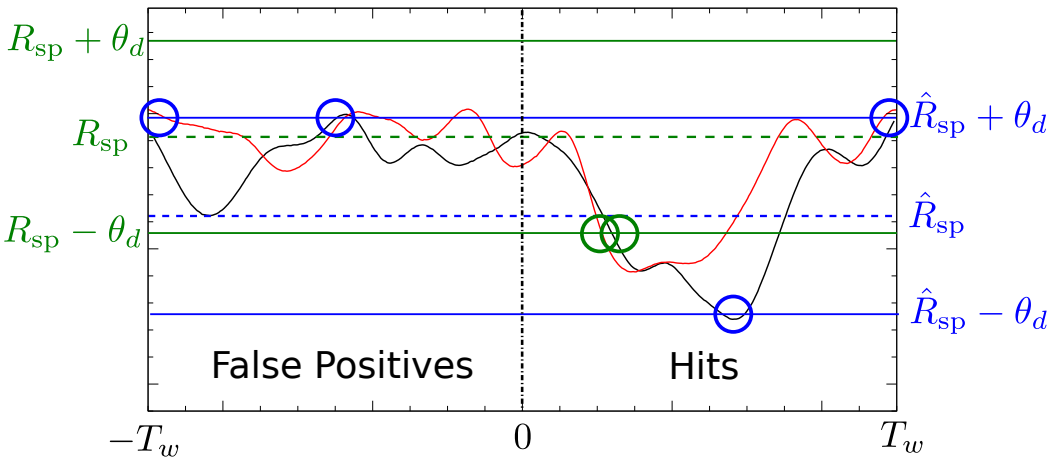

Figure B.1. – **Illustration of how the misalignment of the double-barrier detector can reduce the performance of the detector.** The "green" detector is centered on the true spontaneous mean activity $R_{\rm sp}$, while the "blue" detector is centered on the lower value $\hat{R}_{\rm sp}$, which hinders its ability to detect downward deflections of the readout activity.

According to their definition, all detectors depend on one parameter only, i.e. the position of the threshold. However, in the case of the double-barrier detector, *two* coordinates are required to place the two barriers: besides the value of θ_d, it is necessary to specify the position of the middle point w. The definition prescribes $w = R_{\rm sp}$, which means that the detector must know *exactly* the mean spontaneous activity. In practice, $R_{\rm sp}$ must be either estimated from the data or from the theory. In both cases there are imprecisions. If the theory is used, the discrepancy with the true value stems both from the imprecision in the determination of the network firing rate $r_{\rm sp}$ and from the variability due to the randomness in the construction of the readout set. If the data are used, there are finite-size fluctuations. Regardless of the source, imprecisions in the estimation of $R_{\rm sp}$ have an impact on the performance of the detector, as explained in the following.

Consider the two detectors depicted in fig. B.1: the "green" detector is centered around the actual $R_{\rm sp}$, while the "blue" detector is centered around the value $\hat{R}_{\rm sp}$. Two trials are shown (black and red line). The green detector detects no false positive and two hits (green circles, only first crossing is marked), which corresponds to an effect size of 100%. The blue detector, however, responds with a false positive *and* a hit for both trials, which gives an effect size equal to zero. If θ_d is increased enough to push the upper boundary above the peaks causing the false positives, the lower barrier moves below the trough, so that no hit is recorded either, giving again an effect size of zero.

The general idea behind this slightly caricatural example is that if the detector is misaligned

on a level below the actual mean firing rate, the upper boundary must be raised to keep the false positive rate at 0.25. In this way, the lower boundary moves down even further away from the mean so that the detection of downwards signals is hindered. For the sake of the illustration, the discrepancy between $\hat{R}_{\mathrm{sp}}$ and the true R_{sp} sketched in fig. B.1 is very large, but so is the signal. For a weaker signal, even a modest miscentering has a clear effect on the effect size.

Figure B.2A compares the performance of the three detectors θ_+, θ_-, and θ_d. The double-barrier detector is centered on $w = 2\,\mathrm{Hz}$, which is only slightly above the true ensemble average ($\approx 1.95\,\mathrm{Hz}$). The readout activities considered here are those obtained from the readout set $\mathcal{S}^B$ when an excitatory cell is stimulated and the bias is on the connections to the excitatory cells (as in fig. 3.7A). In this case, the average deflection caused by the stimulation is upwards for $\lambda > \lambda_0 = 0.05$ and downwards for $\lambda < \lambda_0$, which means that the θ_+ detector (upward pointing triangles and solid lines) gives a large positive effect size for $\lambda > \lambda_0$ and a small negative effect size for $\lambda < \lambda_0$. Conversely, the θ_- detector (downward pointing triangles and dashed lines) gives a moderately large positive effect size for $\lambda < \lambda_0$ and a small negative effect size for $\lambda > \lambda_0$.

The double-barrier θ_d detector (plotted with orange circles and lines) yields a significant and mostly positive effect size for most values of the bias different from the unbiased case $\lambda \neq \lambda_0$ and always lies between the two single-barrier detectors. In the range $\lambda < \lambda_0$, the effect is almost the same as for the θ_- detector, while in the range $\lambda > \lambda_0$ the effect for θ_d first sticks to the one for θ_- and then suddenly bends up towards positive values, but the effect size stays well below that obtained by the θ_+ detector.

Simply put, the double-barrier detector can detect both positive and negative deviations from the mean, but at the cost of a smaller effect size in one of the two directions. In the case of fig. B.2A, positive deflections are disadvantaged because the detector is shifted upwards with respect to the true mean so that the upper barrier is further away from potential upwards signals. If it had been the other way around, detection of negative signals would have been hindered, as in the cartoon example of fig. B.1. These observations provide a further reason why in chapter 2 redrawing the network topology was avoided in combination with the double-barrier detector: the variability in the network baseline firing rate due to different network realizations causes the average firing rate for each trial to be sometimes above and sometimes below the ensemble average, thus randomly hampering the detection in one of the two directions.

In fact, even if the detector could be perfectly placed on the exact R_{sp}, two barriers simply detect more false positives than a single barrier, for a given distance from R_{sp}. Hence, the double-barrier detector must increase the distance of the threshold from R_{sp} - compared to the single-barrier detector - to keep the false positive rate at the same level. If the direction of the signal could be known in advance, the non-relevant barrier could be removed to improve the effect.

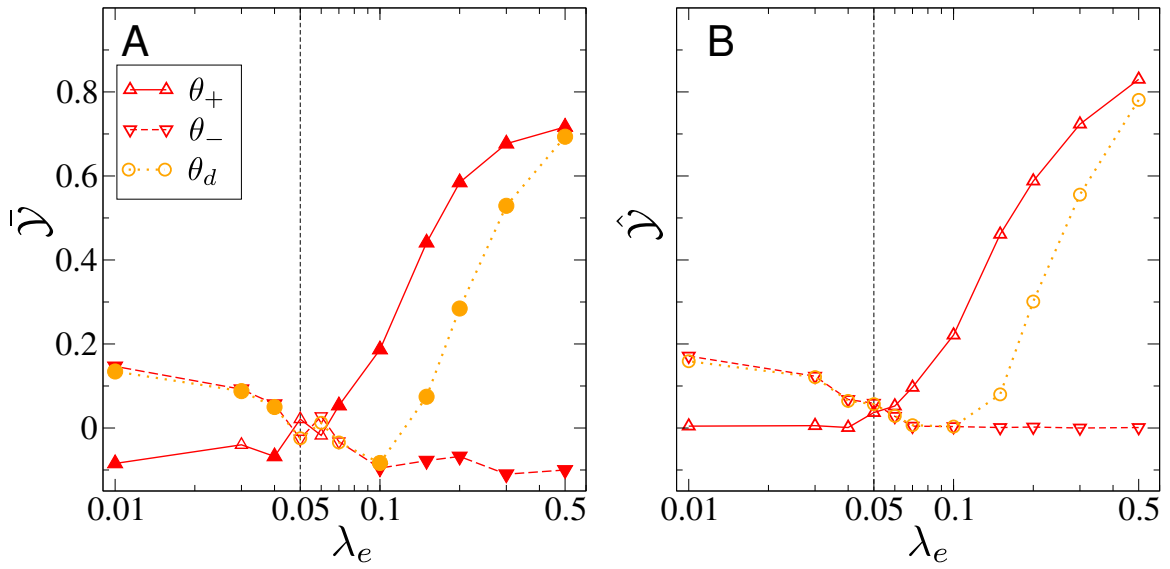

Figure B.2. – **Comparison different detector types considered in this thesis and by Bernardi and Lindner (2017). A**: Effect size according to eq. (B.1) obtained from single-upper-barrier detector θ_+ (red upward pointing triangles and continuous lines), from single-upper-barrier detector θ_- (red downward pointing triangles and dashed lines), and for double-barrier detector θ_d (orange circles and dotted lines). Closed symbols indicate data points significantly different from zero ($p < 0.05$). Data are the same as in fig. 3.7A. **B**: Effect size according to eq. (B.2) for the same data as in **A**. Here, statistical significance is not indicated for reasons explained in the main text.

B.2. Fixed false positive rate vs. optimal threshold

Bernardi and Lindner (2017) used the double-barrier detector in combination with a slightly different definition of the of the effect size: instead of prescribing a fixed false positive rate as in this thesis, they chose the optimal θ_d, that is, the threshold maximizing the effect size

$$\hat{\mathcal{Y}} = \max_{\theta_d} \Big\{ \mathcal{Y}(\theta_d) \Big\} = \max_{\theta_d} \Big\{ \mathcal{CD}(\theta_d) - \mathcal{FP}(\theta_d) \Big\}. \tag{B.2}$$

From this definition, it follows that $0 \leq \hat{\mathcal{Y}} \leq 1$ (the allowed range for the other definition is $-0.25 \leq \bar{\mathcal{Y}} \leq 0.75$) and that $\hat{\mathcal{Y}} \geq \bar{\mathcal{Y}}$. These differences can be seen by comparing fig. B.1A to fig. B.1B, in which the same data are used to recompute the effect size according to eq. (B.2). The qualitative behavior of the two definitions of the effect size is similar and, when $\bar{\mathcal{Y}} > 0$, even quantitative differences are modest. The difference between $\bar{\mathcal{Y}}$ and $\hat{\mathcal{Y}}$ depends on where and on how pronounced the maximum of $\mathcal{Y}(\theta_d)$ is. As observed when discussing the ROC curves in figs. 2.12 and 2.13 on p. 74, the threshold maximizing the effect size (the distance from the ROC curve to the diagonal) corresponds to lower false positive rates for stronger signals and to higher

false positive rates when the signal is weaker and the ROC curve is closer to the diagonal.

One advantage of using the optimal threshold is that it can - by definition - only increase the effect size. However, the enhancement is modest and the definition has two shortcomings: first, the assessment of the statistical significance of data points poses some technical complications, which is discussed in detail below; secondly, it is less consistent with the experimental situation, because it supposes that the threshold can be optimized during the training phase. However, this optimization would require training rats by using repeated single-cell stimulation instead of microstimulation, which turned out to be unsuccessful (Michael Brecht, personal communication). This inconsistency is absent if the other definition of effect size is used because training the detector to react with a prescribed false positive rate requires only the stationary activity and does not require any stimulus.

Statistical significance of the effect size

As already mentioned in chapter 2, optimizing the threshold requires special care in the determination of the statistical significance of data points. In particular, if the same dataset is used both to find the optimal threshold and to calculate p-values with the standard Fisher's test, the results are incorrect. In the remainder of this section, the problem is illustrated by means of a simplified model inspired to the detection theory of section 2.3.3. In this simplified model, the signal is absent and trials used to compute false positive rates and hit rates are obtained under identical conditions. In other words, the null hypothesis is true. In this case, p-values should be uniformly distributed between zero and one (Cox and Hinkley, 1974).

In the detection theory of section 2.3.3, the false positive rate as a function of the threshold θ is given by:

$$\mathcal{FP}(\theta) \approx 1 - p_0^n(\theta), \tag{B.3}$$

where $p_0(\theta)$ is the probability for the readout activity to be inside the range that does not trigger a reaction (i.e. being below θ_+ or above θ_- for the single-barrier detector and between the two barriers for the double-barrier detector) and n is the number draws of the Gaussian variable representing the readout activity. For concreteness, the number of draws is set here to $n = 10$ and $p_0(\theta)$ is computed by applying the θ_+ detector to the data of fig. B.2 with readout $\mathcal{S}^B$, but the following considerations depend neither on the particular value of n nor on the choice of the particular dataset.

In practice, two sets of n samples from the distribution of the stationary activity $R^B(t)$ without stimulus are drawn independently. If at least one sample out of n in the first set is above the detection threshold θ_+, a false positive event is registered. If at least one sample out of n in the second set is above the detection threshold θ_+, a correct detection event is recorded. Because the two sets are drawn independently from the same distribution, this procedure can only produce

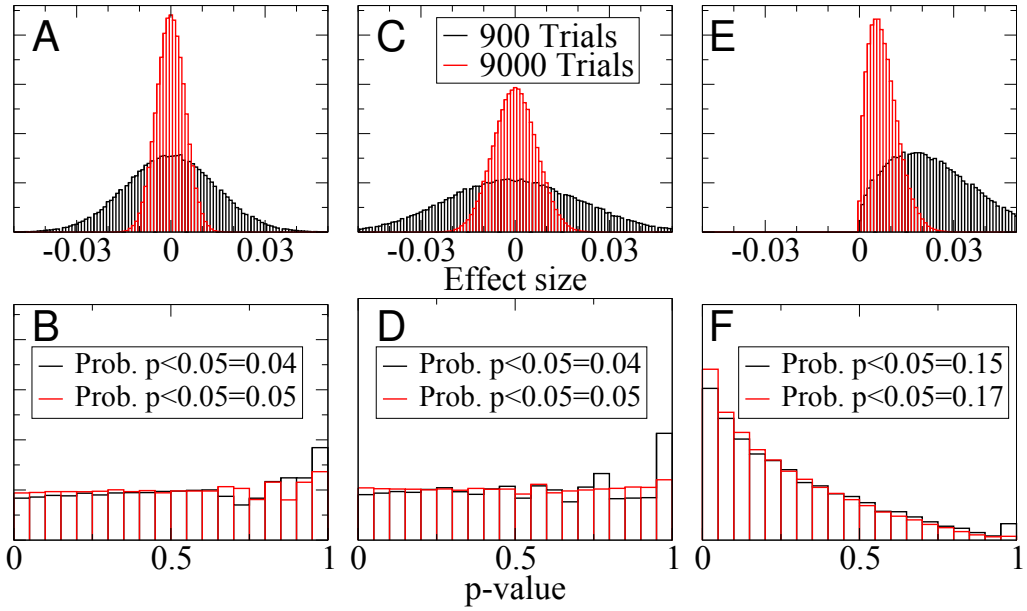

Figure B.3. – **Simplified detection experiment without signal in order to mimic catch trials.** Distribution of effect size (here only due to finite-size fluctuations) and p-value calculated with Fisher's test for three ways of choosing the threshold: fixed threshold (**A**,**B**), threshold corresponding to fixed false positive rate (**C**,**D**), and optimal threshold (**E**,**F**).

a non-zero effect because of random finite-size fluctuations. False positive and correct detection rates are obtained by repeating this random sampling $N_{\text{trials}} = 900$ times - the number of trials used in most simulations of chapters 2 and 3 - and averaging. Subtracting the false positive rate from the correct detection rate defines the effect size, which, because of the finite size of the sample, will not be exactly zero, but randomly distributed. Depending on the procedure used to choose the threshold, the resulting distribution will be different. Three possibilities will be considered here: i) a threshold is fixed arbitrarily *a priori* and used to determine both false positive and correct detection rate; ii) a false positive rate is fixed, the corresponding threshold is determined, then this threshold is used to calculate the correct detection rate (which is the procedure adopted in this thesis); iii) the threshold that maximizes the effect size is selected and employed to determine both false positive and correct detection rate, which is the method employed by Bernardi and Lindner (2017). For each case, the procedure described above was repeated 200000 times. Each repetition of the procedure yields one effect size and one p-value obtained from Fisher's exact test. A histogram of the effect sizes and p-values can be constructed for each of the three cases, as shown in fig. B.3.

If the threshold is fixed beforehand (case i) the effect size has the distribution of the difference of two binomial variables, which, for a large number of trials, it is approximately Gaussian and symmetric (fig. B.3A). Figure B.3B shows that the corresponding p-values are roughly uniformly distributed, although not perfectly because of the finite number of trials. As a matter of fact,

the distribution of p-values is perfectly uniform only for a continuous variable. Because here the number of possible outcomes is discrete, the distribution of p-values is not exactly uniform. Increasing the number of trials by a factor ten (red histograms) renders the histogram of effect sizes narrower and the histogram of p-values flatter.

If a false positive level is fixed and used to determine the threshold (case ii) the histogram is still symmetric (fig. B.3C) but has an increased width, due to the variability in the threshold, which is not fixed as in the previous case. Because the procedure used to select the threshold does not introduce any dependence between false positives and correct detections, p-values are still uniformly distributed as in the case of fixed threshold (fig. B.3D).

Finally, if the threshold is optimized with respect to the effect size (case iii), the effect size can only be positive and the corresponding histogram cannot be symmetric (fig. B.3E). The mean value of the effect size is not zero, and the histogram of p-values is not flat (fig. B.3F). In particular, the probability of $p < 0.05$ is three times larger than 5%, and therefore its value does not express the intended statistical significance. The problem with this procedure is that it introduces a dependence between false positives and correct detections, which violates the null hypothesis.

One conceptually simple - but computationally expensive - way to solve this problem would be to generate two distinct datasets, and use the first one to determine the optimal threshold and the second one to compute effect sizes.

Bibliography

Allen, C. and Stevens, C. F. An evaluation of causes for unreliability of synaptic transmission. *Proc. Natl. Acad. Sci. U.S.A.*, 91(22):10380–10383, 1994.

Alloway, K. D., Zhang, M., and Chakrabarti, S. Septal columns in rodent barrel cortex: functional circuits for modulating whisking behavior. *J. Comp. Neurol.*, 480(3):299–309, 2004.

Almog, M. and Korngreen, A. Is realistic neuronal modeling realistic? *J. Neurophysiol.*, 116(5): 2180–2209, 2016.

Amari, S.-I. Neural theory of association and concept-formation. *Biol. Cybern.*, 26(3):175–185, 1977.

Amit, D. J. and Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex*, 7(3):237–252, 1997a.

Amit, D. J. and Brunel, N. Dynamics of a recurrent network of spiking neurons before and following learning. *Network*, 8(4):373–404, 1997b.

Amitai, Y., Gibson, J. R., Beierlein, M., Patrick, S. L., Ho, A. M., Connors, B. W., and Golomb, D. The spatial dimensions of electrically coupled networks of interneurons in the neocortex. *J. Neurosci.*, 22(10):4142–4152, 2002.

Asanuma, H. and Sakata, H. Functional organization of a cortical efferent system examined with focal depth stimulation in cats. *J. Neurophysiol.*, 30(1):35–54, 1967.

Averbeck, B. B., Latham, P. E., and Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.*, 7(5):358, 2006.

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C. C. H. Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *J. Neurophysiol.*, 107 (11):3116–3134, 2012.

Badel, L., Lefort, S., Brette, R., Petersen, C. C., Gerstner, W., and Richardson, M. J. Dynamic iv curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J. Neurophysiol.*, 99(2):656–666, 2008.

Barlow, H. B. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1(4):371–394, 1972.

Barth, A. L. and Poulet, J. F. A. Experimental evidence for sparse firing in the neocortex. *Trends Neurosci.*, 35:345–355, 2012.

Beierlein, M., Gibson, J. R., and Connors, B. W. A network of electrically coupled interneurons drives synchronized inhibition in neocortex. *Nat. Neurosci.*, 3:904–910, 2000.

Beierlein, M., Gibson, J. R., and Connors, B. W. Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *J. Neurophysiol.*, 90(5):2987–3000, 2003.

Benda, J. and Herz, A. V. A universal model for spike-frequency adaptation. *Neural Comput.*, 15(11):2523–2564, 2003.

Bernardi, D. and Lindner, B. Optimal detection of a localized perturbation in random networks of integrate-and-fire neurons. *Phys. Rev. Lett.*, 118:268301, 2017.

Bernardi, D. and Lindner, B. Detecting single-cell stimulation in a large network of integrate-and-fire neurons. *Phys. Rev. E*, 99(3):032304, 2019.

Bialek and Zee. Understanding the efficiency of human perception. *Phys. Rev. Lett.*, 61:1512–1515, 1988.

Bonifazi, P., Goldin, M., Picardo, M. A., Jorquera, I., Cattani, A., Bianconi, G., Represa, A., Ben-Ari, Y., and Cossart, R. Gabaergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science*, 326(5958):1419–1424, 2009.

Borst, J. G. G. The low synaptic release probability in vivo. *Trends Neurosci.*, 33(6):259–266, 2010.

Brecht, M. and Sakmann, B. Dynamic representation of whisker deflection by synaptic potentials in spiny stellate and pyramidal cells in the barrels and septa of layer 4 rat somatosensory cortex. *J. Physiol.*, 543:49–70, 2002.

Brecht, M., Fee, M. S., Garaschuk, O., Helmchen, F., Margrie, T. W., Svoboda, K., and Osten, P. Novel approaches to monitor and manipulate single neurons in vivo. *J. Neurosci.*, 24:9223–9227, 2004a.

Brecht, M., Schneider, M., Sakmann, B., and Margrie, T. W. Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature*, 427(6976):704–710, 2004b.

Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, 8(3):183–208, 2000.

Brunel, N. and Wang, X. J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.*, 11(1):63–85, 2001.

Brunel, N. Is cortical connectivity optimized for storing information? *Nat. Neurosci.*, 19: 749–755, 2016.

Brunel, N., Chance, F. S., Fourcaud, N., and Abbott, L. Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys. Rev. Lett.*, 86(10):2186, 2001.

Butovas, S. and Schwarz, C. Spatiotemporal effects of microstimulation in rat neocortex: a parametric study using multielectrode recordings. *J. Neurophysiol.*, 90(5):3024–3039, 2003.

Butovas, S., Hormuzdi, S. G., Monyer, H., and Schwarz, C. Effects of electrically coupled inhibitory networks on local neuronal responses to intracortical microstimulation. *J. Neurophysiol.*, 96(3):1227–1236, 2006.

Chakrabarti, S., Zhang, M., and Alloway, K. D. Mi neuronal responses to peripheral whisker stimulation: relationship to neuronal activity in si barrels and septa. *J. Neurophysiol.*, 100 (1):50–63, 2008.

Chklovskii, D. B., Mel, B., and Svoboda, K. Cortical rewiring and information storage. *Nature*, 431(7010):782, 2004.

Cowan, A. I. and Stricker, C. Functional connectivity in layer iv local excitatory circuits of rat somatosensory cortex. *J. Neurophysiol.*, 92(4):2137–2150, 2004.

Cowan, W. M., Südhof, T. C., and Stevens, C. F. *Synapses*. The John Hopkins University Press, Baltimore, 2003.

Cox, D. R. and Hinkley, D. V. *Theoretical statistics*, volume 1. London, Chapman & Hall, 1974.

Cox, D. R. and Lewis, P. A. W. *The Statistical Analysis of Series of Events*. Methuen, London, 1966.

da Costa, N. M. and Martin, K. A. C. Whose cortical column would that be? *Front. Neuroanat.*, 4:16, 2010.

Dayan, P. and Abbott, L. F. *Theoretical neuroscience: computational and mathematical theory of neural systems*. MIT Cambridge: MIT Press, 2001.

de Kock, C. P. J., Bruno, R. M., Spors, H., and Sakmann, B. Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. *J. Physiol.*, 581 (Pt 1):139–154, 2007.

Destexhe, A. and Rudolph-Lilith, M. *Neuronal Noise*. Springer, Boston, USA, 2012.

Diamond, M. E., Von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., and Ahissar, E. 'where'and'what'in the whisker sensorimotor system. *Nat. Rev. Neurosci.*, 9(8):601, 2008.

Doose, J., Doron, G., Brecht, M., and Lindner, B. Noisy juxtacellular stimulation in vivo leads to reliable spiking and reveals high-frequency coding in single neurons. *J. Neurosci.*, 36: 11120–11132, 2016.

Doron, G. *Psychophysical characterization of single neuron stimulation effects in rat barrel cortex*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2012.

Doron, G. and Brecht, M. What single-cell stimulation has told us about neural coding. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 370:20140204, 2015.

Doron, G., von Heimendahl, M., Schlattmann, P., Houweling, A. R., and Brecht, M. Spiking irregularity and frequency modulate the behavioral report of single-neuron stimulation. *Neuron*, 81(3):653–663, 2014.

Douglas, R. J. and Martin, K. A. C. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27(1):419–451, 2004. PMID: 15217339.

Douglas, R. J. and Martin, K. A. C. Mapping the matrix: the ways of neocortex. *Neuron*, 56 (2):226–238, 2007a.

Douglas, R. J. and Martin, K. A. C. Recurrent neuronal circuits in the neocortex. *Curr. Biol.*, 17(13):R496–R500, 2007b.

Droste, F. *Signal transmission in stochastic neuron models with non-white or non-Gaussian noise*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2015.

Droste, F. and Lindner, B. Exact analytical results for integrate-and-fire neurons driven by excitatory shot noise. *J. Comput. Neurosci.*, 43:81–91, 2017a.

Droste, F. and Lindner, B. Exact results for power spectrum and susceptibility of a leaky integrate-and-fire neuron with two-state noise. *Phys. Rev. E*, 95:012411, 2017b.

Dummer, B., Wieland, S., and Lindner, B. Self-consistent determination of the spike-train power spectrum in a neural network with sparse connectivity. *Front. Comput. Neurosci.*, 8:104, 2014.

Feldmeyer, D. Excitatory neuronal connectivity in the barrel cortex. *Front. Neuroanat.*, 6:24, 2012.

Feldmeyer, D., Roth, A., and Sakmann, B. Monosynaptic connections between pairs of spiny stellate cells in layer 4 and pyramidal cells in layer 5a indicate that lemniscal and paralemniscal afferent pathways converge in the infragranular somatosensory cortex. *J. Neurosci.*, 25(13): 3423–3431, 2005.

Feldmeyer, D., Brecht, M., Helmchen, F., Petersen, C. C. H., Poulet, J. F. A., Staiger, J. F., Luhmann, H. J., and Schwarz, C. Barrel cortex function. *Prog. Neurobiol.*, 103:3–27, 2013.

Fisher, R. A. *Statistical methods for research workers*. Biological monographs and manuals. Oliver and Boyd, Edinburgh, 12th edition, 1954.

Flint, A. C., Maisch, U. S., Weishaupt, J. H., Kriegstein, A. R., and Monyer, H. Nr2a subunit expression shortens nmda receptor synaptic currents in developing neocortex. *J. Neurosci.*, 17(7):2469–2476, 1997.

Földiak, P. Forming sparse representations by local anti-hebbian learning. *Biol. Cybern.*, 64(2): 165–170, 1990.

Galarreta, M. and Hestrin, S. A network of fast-spiking cells in the neocortex connected by electrical synapses. *Nature*, 402(6757):72, 1999.

Gardiner, C. W. *Handbook of Stochastic Methods*. Springer-Verlag, Berlin, 1985.

Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F., and Petersen, C. C. H. Unique functional properties of somatostatin-expressing gabaergic neurons in mouse barrel cortex. *Nat. Neurosci.*, 15:607–612, 2012.

Gerstner, W. and Naud, R. How good are neuron models? *Science*, 326:379–380, 2009.

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

Gibson, J. R., Beierlein, M., and Connors, B. W. Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, 402:75–79, 1999.

Gottlieb, J. P. and Keller, A. Intrinsic circuitry and physiological properties of pyramidal neurons in rat barrel cortex. *Exp. Brain Res.*, 115(1):47–60, 1997.

Grill-Spector, K. and Malach, R. The human visual cortex. *Annu. Rev. Neurosci.*, 27:649–677, 2004.

Gross, C. G. Genealogy of the "grandmother cell". *Neuroscientist*, 8:512–518, 2002.

Harris, K. D. and Thiele, A. Cortical state and attention. *Nat. Rev. Neurosci.*, 12(9):509–523, 2011.

Harrison, P. M., Badel, L., Wall, M. J., and Richardson, M. J. Experimentally verified parameter sets for modelling heterogeneous neocortical pyramidal-cell populations. *PLoS Comput. Biol.*, 11(8):e1004165, 2015.

Helias, M., Tetzlaff, T., and Diesmann, M. Echoes in correlated neural systems. *New J. Phys.*, 15:023002, 2013.

Helias, M., Tetzlaff, T., and Diesmann, M. The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol.*, 10(1):1–21, 2014.

Helmstaedter, M., Staiger, J. F., Sakmann, B., and Feldmeyer, D. Efficient recruitment of layer 2/3 interneurons by layer 4 input in single columns of rat somatosensory cortex. *J. Neurosci.*, 28(33):8273–8284, 2008.

Helmstaedter, M., Sakmann, B., and Feldmeyer, D. Neuronal correlates of local, lateral, and translaminar inhibition with reference to cortical columns. *Cereb. Cortex*, 19(4):926–937, 2009a.

Helmstaedter, M., Sakmann, B., and Feldmeyer, D. L2/3 interneuron groups defined by multiparameter analysis of axonal projection, dendritic geometry, and electrical excitability. *Cereb. Cortex*, 19(4):951–962, 2009b.

Herculano-Houzel, S., Ribeiro, P., Campos, L., da Silva, A. V., Torres, L. B., Catania, K. C., and Kaas, J. H. Updated neuronal scaling rules for the brains of glires (rodents/lagomorphs). *Brain. Behav. Evol.*, 78(4):302–314, 2011.

Histed, M. H., Bonin, V., and Reid, R. C. Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron*, 63(4):508–522, 2009.

Hodgkin, A. L. and Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117(4):500–544, 1952.

Holzbecher, A. and Kempter, R. Interneuronal gap junctions increase synchrony and robustness of hippocampal ripple oscillations. *Eur. J. Neurosci.*, 48(12):3446–3465, 2018.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, 79(8):2554–2558, 1982.

Houweling, A. R. and Brecht, M. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451(7174):65–68, 2008.

Houweling, A. R., Doron, G., Voigt, B. C., Herfst, L. J., and Brecht, M. Nanostimulation: manipulation of single neuron activity by juxtacellular current injection. *J. Neurophysiol.*, 103(3):1696–1704, 2010.

Hu, Y., Zylberberg, J., and Shea-Brown, E. The sign rule and beyond: boundary effects, flexibility, and noise correlations in neural population codes. *PLoS Comput. Biol.*, 10(2): e1003469, 2014.

Jonas, P., Bischofberger, J., and Sandkühler, J. Corelease of two fast neurotransmitters at a central synapse. *Science*, 281(5375):419–424, 1998.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. *Principles of Neural Science*. McGraw-Hill New York, 4th edition, 2000.

Kapfer, C., Glickfeld, L. L., Atallah, B. V., and Scanziani, M. Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex. *Nat. Neurosci.*, 10(6):743–753, 2007.

Keener, J. and Sneyd, J. *Mathematical Physiology*. Springer-Verlag New York, 2009.

Knutsen, P. M. and Ahissar, E. Orthogonal coding of object location. *Trends Neurosci.*, 32(2): 101–109, 2009.

Koelbl, C., Helmstaedter, M., Lübke, J., and Feldmeyer, D. A barrel-related interneuron in layer 4 of rat somatosensory cortex with a high intrabarrel connectivity. *Cereb. Cortex*, 25(3): 713–725, 2015.

Korbo, L., Pakkenberg, B., Ladefoged, O., Gundersen, H. J. G., Arlien-Søborg, P., and Pakkenberg, H. An efficient method for estimating the total number of neurons in rat brain cortex. *J. Neurosci. Methods*, 31(2):93 – 100, 1990.

Kost, J. T. and McDermott, M. P. Combining dependent p-values. *Stat. Probabil. Lett.*, 60: 183–190, 2002.

Kruscha, A. and Lindner, B. Spike-count distribution in a neuronal population under weak common stimulation. *Phys. Rev. E*, 92:052817, 2015.

Kruscha, A. and Lindner, B. Partial synchronous output of a neuronal population under weak common noise: analytical approaches to the correlation statistics. *Phys. Rev. E*, 94:022422, 2016.

Kwan, A. C. and Dan, Y. Dissection of cortical microcircuits by single-neuron stimulation in vivo. *Curr. Biol.*, 22(16):1459–1467, 2012.

Lánskỳ, P. and Lanska, V. Diffusion approximation of the neuronal model with synaptic reversal potentials. *Biol. Cybern.*, 56(1):19–26, 1987.

Larkum, M. A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.*, 36(3):141–151, 2013.

Ledoux, E. and Brunel, N. Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. *Front. Comput. Neurosci.*, 5:25, 2011.

Lefort, S. and Petersen, C. C. H. Layer-dependent short-term synaptic plasticity between excitatory neurons in the c2 barrel column of mouse primary somatosensory cortex. *Cereb. Cortex*, 27:3869–3878, 2017.

Lefort, S., Tomm, C., Sarria, J.-C. F., and Petersen, C. C. H. The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61(2):301–316, 2009.

Lerchner, A., Ursta, C., Hertz, J., Ahmadi, M., Ruffiot, P., and Enemark, S. Response variability in balanced cortical networks. *Neural Comput.*, 18(3):634–659, 2006.

Lewis, T. J. and Rinzel, J. Dynamics of spiking neurons connected by both inhibitory and electrical coupling. *J. Comput. Neurosci.*, 14(3):283–309, 2003.

Li, C.-y. T., Poo, M.-m., and Dan, Y. Burst spiking of a single cortical neuron modifies global brain state. *Science*, 324(5927):643–646, 2009.

Lindner, B. Neural noise and neural signals - spontaneous activity and information transmission in models of single nerve cells, 2013. Lecture Notes, Summer Semester 2013, Bernstein Center for Computational Neuroscience, Berlin.

Lindner, B. and Schimansky-Geier, L. Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys. Rev. Lett.*, 86:2934–2937, 2001.

Lindner, B., Doiron, B., and Longtin, A. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys. Rev. E.*, 72:061919, 2005.

Lindner, B. *Coherence and stochastic resonance in nonlinear dynamical systems.* Logos-Verlag, 2002.

Lindner, B. Superposition of many independent spike trains is generally not a poisson process. *Phys. Rev. E*, 73(2):022901, 2006.

Litwin-Kumar, A. and Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.*, 15(11):1498–1505, 2012.

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123–127, 2010.

Lübke, J., Markram, H., Frotscher, M., and Sakmann, B. Frequency and dendritic distribution of autapses established by layer 5 pyramidal neurons in the developing rat neocortex: Comparison with synaptic innervation of adjacent neurons of the same class. *J. Neurosci.*, 16(10): 3209–3218, 1996.

Luccioli, S., Ben-Jacob, E., Barzilai, A., Bonifazi, P., and Torcini, A. Clique of functional hubs orchestrates population bursts in developmentally regulated neural networks. *PLoS Comput. Biol.*, 10(9):e1003823, 2014.

Markram, H., Wang, Y., and Tsodyks, M. Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. U.S.A.*, 95:5323–5328, 1998.

Martens, M. B., Houweling, A. R., and Tiesinga, P. H. E. Anti-correlations in the degree distribution increase stimulus detection performance in noisy spiking neural networks. *J. Comput. Neurosci.*, 42:87–106, 2017.

Meyer, H. S., Wimmer, V. C., Oberlaender, M., de Kock, C. P. J., Sakmann, B., and Helmstaedter, M. Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. *Cereb. Cortex*, 20:2277–2286, 2010.

Meyer, H. S., Schwarz, D., Wimmer, V. C., Schmitt, A. C., Kerr, J. N. D., Sakmann, B., and Helmstaedter, M. Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5a. *Proc. Natl. Acad. Sci. U.S.A.*, 108(40):16807–16812, 2011.

Middleton, J. W., Omar, C., Doiron, B., and Simons, D. J. Neural correlation is stimulus modulated by feedforward inhibitory circuitry. *J. Neurosci.*, 32:506–518, 2012.

Miyashita, E., Keller, A., and Asanuma, H. Input-output organization of the rat vibrissal motor cortex. *Exp. Brain Res.*, 99(2):223–232, 1994.

Monteforte, M. and Wolf, F. Dynamical entropy production in spiking neuron networks in the balanced state. *Phys. Rev. Lett.*, 105(26):268104, 2010.

Mountcastle, V. B. The columnar organization of the neocortex. *Brain*, 120(4):701–722, 1997.

Newman, M. The structure and function of complex networks. *SIAM Rev.*, 45:167–256, 2003.

Olshausen, B. A. and Field, D. J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.*, 14 (4):481–487, 2004.

Ostojic, S., Brunel, N., and Hakim, V. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.*, 29:10234, 2009a.

Ostojic, S. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci.*, 17(4):594–600, 2014.

Ostojic, S., Brunel, N., and Hakim, V. Synchronization properties of networks of electrically coupled neurons in the presence of noise and heterogeneities. *J. Comput. Neurosci.*, 26(3): 369–392, 2009b.

Overstreet, C. K., Klein, J. D., and Tillery, S. I. H. Computational modeling of direct neuronal recruitment during intracortical microstimulation in somatosensory cortex. *J. Neural Eng.*, 10(6):066016, 2013.

Packer, A. M. and Yuste, R. Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *J. Neurosci.*, 31:13260–13271, 2011.

Paxinos, G. and Watson, C. *The rat brain in stereotaxic coordinates: hard cover edition*. Elsevier, 2006.

Pehlevan, C. and Sompolinsky, H. Selectivity and sparseness in randomly connected balanced networks. *PLoS One*, 9(2):e89992, 2014.

Pena, R. F., Vellmer, S., Bernardi, D., Roque, A. C., and Lindner, B. Self-consistent scheme for spike-train power spectra in heterogeneous sparse networks. *Front. Comput. Neurosci.*, 12:9, 2018.

Petersen, C. C. H., Grinvald, A., and Sakmann, B. Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. *J. Neurosci.*, 23(4):1298–1309, 2003.

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., and Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.*, 16 (8):1068–1076, 2013.

Pinault, D. Golgi-like labeling of a single neuron recorded extracellularly. *Neurosci. Lett.*, 170 (2):255–260, 1994.

Poulet, J. F. A., Fernandez, L. M. J., Crochet, S., and Petersen, C. C. H. Thalamic control of cortical states. *Nat. Neurosci.*, 15:370–372, 2012.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102, 2005.

Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.*, 10(10):724, 2009.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. The asynchronous state in cortical circuits. *Science*, 327(5965):587–590, 2010.

Ricciardi, L. M. and Sacerdote, L. The ornstein-uhlenbeck process as a model for neuronal activity. i. mean and variance of the firing time. *Biol. Cybern.*, 35:1–9, 1979.

Richardson, M. J. E. and Swarbrick, R. Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise. *Phys. Rev. Lett.*, 105(17):178102, 2010.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. *Spikes: Exploring the neural code*. MIT Press, Cambridge, Massachusetts, 1999.

Rockel, A. J., Hiorns, R. W., and Powell, T. P. The basic uniformity in structure of the neocortex. *Brain*, 103(2):221–244, 1980.

Rolls, E. T. Cortical coding. *Lang. Cogn. Neurosci.*, 32(3):316–329, 2017.

Rolls, E. T. and Deco, G. *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford University Press, 2010.

Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. The spatial structure of correlated neuronal variability. *Nat. Neurosci.*, 20(1):107–114, 2017.

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. On the distribution of firing rates in networks of cortical neurons. *J. Neurosci.*, 31(45):16217–16226, 2011.

Salzman, C. D. and Newsome, W. T. Neural mechanisms for forming a perceptual decision. *Science*, 264(5156):231–237, 1994.

Salzman, C. D., Britten, K. H., and Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346(6280):174–177, 1990.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

Schnepel, P., Kumar, A., Zohar, M., Aertsen, A., and Boucsein, C. Physiology and impact of horizontal connections in rat neocortex. *Cereb. Cortex*, 25:3818–3835, 2014.

Schoonover, C. E., Tapia, J.-C., Schilling, V. C., Wimmer, V., Blazeski, R., Zhang, W., Mason, C. A., and Bruno, R. M. Comparative strength and dendritic organization of thalamocortical and corticocortical synapses onto excitatory layer 4 neurons. *J. Neurosci.*, 34(20): 6746–6758, 2014.

Schubert, D., Kötter, R., and Staiger, J. F. Mapping functional connectivity in barrel-related columns reveals layer- and cell type-specific microcircuits. *Brain Struct. Funct.*, 212(2):107–119, 2007.

Schwalger, T. and Lindner, B. Patterns of interval correlations in neural oscillators with adaptation. *Front. Comput. Neurosci.*, 7:164, 2013.

Schwalger, T., Deger, M., and Gerstner, W. Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. *PLoS Comput. Biol.*, 13(4):e1005507, 2017.

Sejnowski, T. J. On the stochastic dynamics of neuronal interaction. *Biol. Cybern.*, 22(4): 203–211, 1976a.

Sejnowski, T. J. On global properties of neuronal interaction. *Biol. Cybern.*, 22(2):85–95, 1976b.

Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M., and Noble-Haeusslein, L. J. Brain development in rodents and humans: identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.*, 106-107:1–16, 2013.

Shadlen, M. N. and Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.*, 18(10):3870–3896, 1998.

Silberberg, G. and Markram, H. Disynaptic inhibition between neocortical pyramidal cells mediated by martinotti cells. *Neuron*, 53(5):735–746, 2007.

Sompolinsky, Crisanti, and Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61 (3):259–262, 1988.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.*, 3(3):e68, 2005.

Stocks, N. G. Suprathreshold stochastic resonance in multilevel threshold systems. *Phys. Rev. Lett.*, 84:2310–2313, 2000.

Strata, P. and Harvey, R. Dale' s principle. *Brain Res. Bull.*, 50(5-6):349–350, 1999.

Stratonovich, R. *Topics on the Theory of Random Noise. Vol 1.* Gordon and Breach, New York, 1963.

Sun, Q.-Q., Huguenard, J. R., and Prince, D. A. Barrel cortex microcircuits: thalamocortical feedforward inhibition in spiny stellate cells is mediated by a small number of fast-spiking interneurons. *J. Neurosci.*, 26(4):1219–1230, 2006.

Tamás, G., Buhl, E. H., Lörincz, A., and Somogyi, P. Proximally targeted gabaergic synapses and gap junctions synchronize cortical interneurons. *Nat. Neurosci.*, 3(4):366, 2000.

Tanke, N., Borst, J. G. G., and Houweling, A. R. Single-cell stimulation in barrel cortex influences psychophysical detection performance. *J. Neurosci.*, 38:2057–2068, 2018.

Tehovnik, E. J. Electrical stimulation of neural tissue to evoke behavioral responses. *J. Neurosci. Methods*, 65(1):1–17, 1996.

Tremblay, R., Lee, S., and Rudy, B. Gabaergic interneurons in the neocortex: From cellular properties to circuits. *Neuron*, 91:260–292, 2016.

Treves, A. Are spin-glass effects relevant to understanding realistic auto-associative networks? *J. Phys. A: Math. Gen.*, 24(11):2645, 1991.

Treves, A. Mean-field analysis of neuronal spike dynamics. *Network*, 4(3):259–284, 1993.

Trousdale, J., Hu, Y., Shea-Brown, E., and Josić, K. Impact of network structure and cellular response on spike time correlations. *PLoS Comput. Biol.*, 8(3):e1002408, 2012.

Tsodyks, M., Pawelzik, K., and Markram, H. Neural networks with dynamic synapses. *Neural Comput.*, 10:821–835, 1998.

Tsodyks, M. V. and Markram, H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. U.S.A.*, 94:719–723, 1997.

van Meegen, A. and Lindner, B. Self-consistent correlations of randomly coupled rotators in the asynchronous state. *Phys. Rev. Lett.*, 121:258302, 2018.

van Vreeswijk, C. and Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

van Vreeswijk, C. and Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural Comput.*, 10(6):1321–1371, 1998.

Vasquez, J. C., Houweling, A. R., and Tiesinga, P. Simultaneous stability and sensitivity in model cortical networks is achieved through anti-correlations between the in- and out-degree of connectivity. *Front. Comput. Neurosci.*, 7:156, 2013.

Vilela, R. D. and Lindner, B. Comparative study of different integrate-and-fire neurons: spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys. Rev. E*, 80: 031909, 2009.

Vincent, S. The function of the vibrissae in the behavior of the white rat. *Behavior Mon.*, 1: 1–82, 1912.

Vogels, T. and Abbott, L. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat. Neurosci.*, 12(4):483–491, 2009.

Voigt, B. C., Brecht, M., and Houweling, A. R. Behavioral detectability of single-cell stimulation in the ventral posterior medial nucleus of the thalamus. *J. Neurosci.*, 28(47):12362–12367, 2008.

Voronenko, S. O. and Lindner, B. Weakly nonlinear response of noisy neurons. *New J. Phys.*, 19(3):033038, 2017.

White, J. A., Rubinstein, J. T., and Kay, A. R. Channel noise in neurons. *Trends Neurosci.*, 23 (3):131–137, 2000.

Wieland, S., Bernardi, D., Schwalger, T., and Lindner, B. Slow fluctuations in recurrent networks of spiking neurons. *Phys. Rev. E*, 92(4):040901, 2015.

Wolfe, J., Houweling, A. R., and Brecht, M. Sparse and powerful cortical spikes. *Curr. Opin. Neurobiol.*, 20(3):306–312, 2010.

Woolsey, T. A. and Van der Loos, H. The structural organization of layer iv in the somatosensory region (si) of mouse cerebral cortex. the description of a cortical field composed of discrete cytoarchitectonic units. *Brain Res.*, 17:205–242, 1970.

Acknowledgments

In the list of people to whom I would like to express my gratitude, my supervisor, Prof. Benjamin Lindner well deserves the first place. I would like to thank him for finding the many hours he spent answering my questions, giving me plenty of new ideas, finding my mistakes, sitting at a desk with me trying to teach me how analytical calculations are done, and encouraging me when things did not work. His commitment well exceeds that of the average supervisor and made me realize that the German word *Doktorvater*, which used to sound a bit funny to me, makes sense.

In the time spent in the research group "Theory of Complex Systems and Neurophysics", I was fortunate to have always been surrounded by friendly and helpful colleagues. I am particularly thankful to all my past and present fellow PhD students Alexandra, Felix, Florian, Greg, Jens, Rinaldo, Sergej, Sebastian, and Sven for their collaborative and humorous attitude. A particular mention deserves Felix Droste for the enormous number of questions he patiently answered over the years and for the huge amount of time I saved thanks to his advice during the development of the simulation software used for this thesis.

I greatly appreciate the fact that Prof. Michael Brecht and Dr. Guy Doron granted me full access to their experimental data. I would like to acknowledge Guy Doron's readiness in answering so many questions about the experimental procedures and to thank him for inviting me to observe one experimental session.

I would like to thank all the people who contributed to creating and running the graduate program GRK 1589/2 "Sensory Computation in Neural Systems", from which I received not only financial support, but also a privileged access channel to other research topics. I would like to give a special recognition to Nikola Schrenk, Robert Martin, and Margret Franke for combining efficiency in administrative matters with friendliness and a good mood. The meetings with the members of my GRK-PhD committee, Prof. Michael Brecht and Prof. Richard Kempter, provided me with helpful advice and constructive feedback, which I would like to acknowledge.

I would like to thank Rob Scheid, Greg Knoll, and Paolo Bernardi for the time and effort they put into carefully proof reading parts of this thesis.

I am grateful to my parents for their continued support throughout all phases of my education and for eventually stopping asking whether the thesis was done. My final thanks goes to all the people who had to bear my bad mood in the difficult phases, in particular Paolo, Rita, and Marianna. I am sure that the single-cell stimulation of the neuron in my sensory cortex that encodes them would be very easy to detect.

Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 17. Juni 2019 Davide Bernardi